

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 3, 2003, 15:36:12 ; Search time 84 Seconds  
(without alignments)  
1219.112 Million cell updates/sec

Title: CAC80065  
Perfect score: 2576  
Sequence: 1 MSALVLRCAVAVSIRGSSCR.....HELLHHTYVAGVEKDEKK 497

Scoring table:  
BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues  
Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database :  
1: SP archaea:\*  
2: SP bacteria:\*  
3: SP fungi:\*  
4: SP human:\*  
5: SP invertebrate:\*  
6: SP mammal:\*  
7: SP mhc:\*  
8: SP organelle:\*  
9: SP phage:\*  
10: SP plant:\*  
11: SP rodent:\*  
12: SP virus:\*  
13: SP vertebrate:\*  
14: SP unclassified:\*  
15: SP virus:\*  
16: SP bacteriophage:\*  
17: SP archaea:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2576	100.0	497	10 Q9FYU1	Q9FYU1 chlamydomon
2	1702.5	66.1	505	10 Q8VZ20	Q8VZ20 chlamydomon
3	1409	54.7	436	10 Q8VX03	Q8VX03 chlamydomon
4	1384.5	53.7	449	10 Q9AR66	Q9AR66 chlamydomon
5	1196	46.4	403	10 Q9AU60	Q9AU60 chlamydomon
6	1010.5	39.2	608	16 Q9WY44	Q9WY44 thermotoga
7	935.5	36.3	581	16 Q8BCE8	Q8BCE8 thermotoga
8	903.5	35.1	579	2 Q9XC55	Q9XC55 clostridium
9	893	34.7	555	3 Q8TGC3	Q8TGC3 pitomycetes s
10	880.5	34.2	636	3 Q8TFP2	Q8TFP2 neocallimastix
11	877	34.0	585	2 Q46508	Q46508 desulfotomobacter
12	874	32.0	1206	5 Q96948	Q96948 neocallimastix
13	816.5	31.7	578	2 Q93SF7	Q93SF7 eubacterium
14	805.5	31.3	589	2 Q9GTP2	Q9GTP2 trichomonas
15	794.5	30.8	606	2 Q46606	Q46606 desulfotomobacter
16	778.5	30.2	484	2 Q9RCN3	Q9RCN3 megasphaera

17	741	28.8	582	16 Q59262	Q59262 clostridium
18	738.5	28.7	468	5 Q27094	Q27094 trichomonas
19	737.5	28.6	574	2 Q59261	Q59261 clostridium
20	732.5	28.4	572	2 Q92NE4	Q92NE4 clostridium
21	730.5	28.4	572	16 Q9RH08	Q9RH08 clostridium
22	659	25.6	421	2 Q08311	Q08311 desulfotomobacter
23	652	25.3	449	5 Q27096	Q27096 trichomonas
24	590.5	22.9	421	2 Q9AM36	Q9AM36 desulfotomobacter
25	555	21.5	467	5 Q9GTP1	Q9GTP1 spirochaeta
26	512.5	19.9	645	16 Q52683	Q52683 thermotoga
27	504.5	19.6	464	5 Q9GTP3	Q9GTP3 entamoeba h
28	504.5	19.6	468	5 Q9GTP3	Q9GTP3 entamoeba h
29	492.5	19.1	476	11 Q9D320	Q9D320 mus musculus
30	489.5	19.0	476	11 Q9C856	Q9C856 mus musculus
31	478	18.6	474	10 Q9AC16	Q9AC16 arabidopsis
32	465.5	18.1	369	5 Q8TSG1	Q8TSG1 trichomonas
33	461.5	17.9	476	10 Q8W303	Q8W303 oryza sativa
34	459	17.8	479	10 Q93YF9	Q93YF9 medicago tr
35	457	17.7	456	5 Q9TY15	Q9TY15 neocallimastix
36	456.5	17.7	476	4 Q9H6Q4	Q9H6Q4 homo sapien
37	456.5	17.7	525	4 Q96S10	Q96S10 homo sapien
38	451	17.5	369	5 Q8TGP9	Q8TGP9 trichomonas
39	446.5	17.3	369	5 Q8TSG2	Q8TSG2 trichomonas
40	445	17.3	474	5 Q9BKJ3	Q9BKJ3 glardia lam
41	437.5	17.0	374	4 Q9H6Q8	Q9H6Q8 homo sapien
42	390.5	15.2	462	11 Q9CYQ7	Q9CYQ7 mus musculus
43	388.5	15.1	477	5 Q8SY57	Q8SY57 drosophila
44	367	14.2	457	5 Q9N392	Q9N392 caenorhabditis
45	361.5	14.0	456	4 Q9UHQ1	Q9UHQ1 homo sapien

## ALIGNMENTS

RESULT 1  
ID Q9FYU1 PRELIMINARY; PRT; 497 AA.  
AC Q9FYU1;  
DT 01-MAR-2001 (TREMBLrel. 16, last sequence update)  
DT 01-MAR-2001 (TREMBLrel. 20, last annotation update)  
DE Fe-hydrogenase precursor (EC 1.18.99.1) (Iron-hydrogenase HydA).  
GN HYD1 OR HYD4.  
OS Chlamydomonas reinhardtii.  
OC Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;  
OC Chlamydomonadaceae; Chlamydomonas.  
OX NCBI\_TaxID=3055;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Kaminski A.U., Happe T.;  
RC "Isolation and characterization of the hydA gene encoding the Fe-hydrogenase of Chlamydomonas reinhardtii.";  
RT Submitted (JUL-2000) to the EMBL/Genbank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Kaminski A.U., Happe T.;  
RC "Two putative Fe-only hydrogenases cloned from Chlamydomonas reinhardtii are coexpressed in cells undergoing anaerobiosis.";  
RT Submitted (SEP-2001) to the EMBL/Genbank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RA Chiriac M.L.;  
RC "Two putative Fe-only hydrogenases cloned from Chlamydomonas reinhardtii are coexpressed in cells undergoing anaerobiosis.";  
RT Submitted (SEP-2001) to the EMBL/Genbank/DBJ databases.  
DR EMBL; AF288201; RAG00591.1; -;  
DR EMBL; AF012098; CAC80065.1; -;  
DR EMBL; AY055755; AAL23572.1; -;  
DR HSP; P29166; 1FER.

DR InterPro: IPR004108; Fe\_hyd\_19\_C.  
 DR InterPro: IPR003149; Fe\_hyd\_SSU.  
 DR Pfam: PF02906; Fe\_hyd\_19\_C; 1.  
 DR Pfam: PF02256; Fe\_hyd\_SSU; 1.  
 DR Oxidoreductase; Transalt peptide.  
 FT TRANSIT 1  
 FT CHAIN 57  
 SQ SEQUENCE 497 AA; 53112 MW; 2E618A259E6572F4 CRC64;

Query Match 100.0%; Score 2576; DB 10; Length 497;  
 Best Local Similarity 100.0%; Pred. NO. 2e-175;  
 Matches 497; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSALVLRPCAAVSTRSSCAROYAPRAPLAATVRAVLAATLEAPARRLGNVACAAAPA 60  
 DB 1 MSALVLRPCAAVSTRSSCAROYAPRAPLAATVRAVLAATLEAPARRLGNVACAAAPA 60  
 QY 61 AEAFLSHVQOALAEALPKDDPTKHKYCVAPAVRAVAIAETGLAAGATTPKOLAEGLR 120  
 DB 61 AEAFLSHVQOALAEALPKDDPTKHKYCVAPAVRAVAIAETGLAAGATTPKOLAEGLR 120  
 QY 121 RLGEDEVDTLFGADLTMEGSELRLTEHLLEAHPSDEPLPMFTSCCPGWTAMLEKS 180  
 DB 121 RLGEDEVDTLFGADLTMEGSELRLTEHLLEAHPSDEPLPMFTSCCPGWTAMLEKS 180  
 QY 181 YPDLIPYVSSCKSFQMLAAVNSYLAEEKGIAPKDMVNSIMPCTRKQSEADRDWFCVD 240  
 DB 181 YPDLIPYVSSCKSFQMLAAVNSYLAEEKGIAPKDMVNSIMPCTRKQSEADRDWFCVD 240  
 QY 241 ADPLRLDHLVITVELGNIFKRGITLAEPLPEGEMDPMGVSGAGVLEGTGGVMEA 300  
 DB 241 ADPLRLDHLVITVELGNIFKRGITLAEPLPEGEMDPMGVSGAGVLEGTGGVMEA 300  
 QY 301 LRTAYELFTGTPRLSLSEVRGMDGKETNTMVPAPGSKFEELKRAAARAAEAAG 360  
 DB 301 LRTAYELFTGTPRLSLSEVRGMDGKETNTMVPAPGSKFEELKRAAARAAEAAG 360  
 QY 361 TPGPLAMDGAGFTSEDRGGITLRAVANGIGNAKKLITKMOAGEAKYDFEIMACPA 420  
 DB 361 TPGPLAMDGAGFTSEDRGGITLRAVANGIGNAKKLITKMOAGEAKYDFEIMACPA 420  
 QY 421 CVGGGQPRSTDKATLQKRAALYNDEKSTLRSHENPSIRELYDTYLGEPGLGHAHEL 480  
 DB 421 CVGGGQPRSTDKATLQKRAALYNDEKSTLRSHENPSIRELYDTYLGEPGLGHAHEL 480  
 QY 481 LHTHYVAGVEERDEK 497  
 DB 481 LHTHYVAGVEERDEK 497

## RESULT 2

Q8VZ20 PRELIMINARY; PRT; 505 AA.  
 AC Q8VZ20;  
 DT 01-MAR-2002 (T-EMBLrel. 20, Created)  
 DT 01-MAR-2002 (T-EMBLrel. 20, Last sequence update)  
 DT 01-JUN-2002 (T-EMBLrel. 21, Last annotation update)  
 DE Iron-hydrogenase HydB.  
 GN HYDB.  
 OS Chlamydomonas reinhardtii.  
 OC Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;  
 OC Chlamydomonadaceae; Chlamydomonas.  
 OK NCBI\_TaxID=3035;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=21GR;  
 RA Forestier M., Zhang L., Plummer S., Ahmann D., Seibert M.,  
 RA Chlarard M.L.;  
 RT "Two putative Fe-only hydrogenases cloned from Chlamydomonas  
 RT reinhardtii are coexpressed in cells undergoing anaerobiosis";  
 RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AY055756; AL23573.1; -;  
 DR InterPro: IPR004108; Fe\_hyd\_19\_C.

DR InterPro: IPR003149; Fe\_hyd\_SSU.  
 DR Pfam: PF02906; Fe\_hyd\_19\_C; 1.  
 DR Pfam: PF02256; Fe\_hyd\_SSU; 1.  
 DR Oxidoreductase; Transalt peptide.  
 FT TRANSIT 1  
 FT CHAIN 57  
 SQ SEQUENCE 505 AA; 53121 MW; 3A1948749B033E9A CRC64;

Query Match 66.1%; Score 1702.5; DB 10; Length 505;  
 Best Local Similarity 67.7%; Pred. No. 3.7e-113;  
 Matches 331; Conservative 57; Mismatches 84; Indels 17; Gaps 5;

QY 18 SCAROVAPRAPLAATVRAVLA-----TLEAPARRLGNVACAAAPAEAPLSHQQA 71  
 DB 16 ACARRTNAPAPAPAVPCLPSRAGKFFNLQKVPSSQASGSTRIVAAATADVPHWKL 75  
 QY 72 LAELAPKDDPTKHKYCVAPAVRAVAIAETGLAAGATTPKOLAEGLR 131  
 DB 72 LAELAPKDDPTKHKYCVAPAVRAVAIAETGLAAGATTPKOLAEGLR 131  
 QY 132 FGADLTMEGSELRLTEHLLEAHPSDEPLPMFTSCCPGWTAMLEKSPDLIPYVSSC 191  
 DB 135 FAADLTMEGSELRLTEHLLEAHPSDEPLPMFTSCCPGWTAMLEKSPDLIPYVSSC 194  
 QY 192 KSPQMLAAVNSYLAEEKGIAPKDMVNSIMPCTRKQSEADRDWFCVDADPTLRDLDV 251  
 DB 195 KSPQMLAAVNSYLAEEKGIAPKDMVNSIMPCTRKQSEADRDWFCVDADPTLRDLDV 253  
 QY 252 ITTVELGNIFKRGITLAEPLPEGEMDPMGVSGAGVLEGTGGVMEAALRTAYELFTG 311  
 DB 254 ITTVELGNIFKRGITLAEPLPEGEMDPMGVSGAGVLEGTGGVMEAALRTAYELFTG 313  
 QY 312 PLRLSLSEVRGMDGKETNTMVPAPGSKFEELKRAAARAAEAAG-TP 362  
 DB 314 PLRLSLSEVRGMDGKETNTMVPAPGSKFEELKRAAARAAEAAG-TP 373  
 QY 363 GPLAMGAGFTSEDRGGITLRAVANGIGNAKKLITKMOAGEAKYDFEIMACPA 422  
 DB 374 GPLAMGAGFTSEDRGGITLRAVANGIGNAKKLITKMOAGEAKYDFEIMACPA 433  
 QY 423 GGGGQPRSTDKATLQKRAALYNDEKSTLRSHENPSIRELYDTYLGEPGLGHAHEL 482  
 DB 434 GGGGQPRSTDKATLQKRAALYNDEKSTLRSHENPSIRELYDTYLGEPGLGHAHEL 493  
 QY 483 THYVAGVE 491  
 DB 494 THYVAGVE 502

## RESULT 3

Q8VX03 PRELIMINARY; PRT; 436 AA.  
 AC Q8VX03;  
 DT 01-MAR-2002 (T-EMBLrel. 20, Created)  
 DT 01-MAR-2002 (T-EMBLrel. 20, Last sequence update)  
 DT 01-JUN-2002 (T-EMBLrel. 21, Last annotation update)  
 DE Fe-hydrogenase precursor (EC 1.18.99.1).  
 GN HYDA.  
 OS Chlorella fusca.  
 OC Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Chlorococcales;  
 OC Scenedesmeaceae; Scenedesmus.  
 OK NCBI\_TaxID=3073;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Happe T.;  
 RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=21GR;  
 RA Winkler M.;  
 RT "The Fe-hydrogenase of Chlorella fusca";  
 RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AJ298228; CAC83291.1; -;  
 DR InterPro: IPR004108; Fe\_hyd\_19\_C.  
 DR InterPro: IPR003149; Fe\_hyd\_SSU.  
 DR Pfam: PF02906; Fe\_hyd\_19\_C; 1.  
 DR Pfam: PF02256; Fe\_hyd\_SSU; 1.

KW Transit peptide: Oxidoreductase.  
 FT TRANSIT 1 21 POTENTIAL.  
 FT CHAIN 22 436 FE-HYDROENASE.  
 SQ SEQUENCE 436 AA; 47318 MW; BF7299732PFI984 CRC64;

Query Match 54.7%; Score 1409; DB 10; Length 436;  
 Best Local Similarity 59.6%; Pred. No. 2.5e-92;  
 Matches 272; Conservative 56; Mismatches 89; Indels 39; Gaps 4;

QY 46 ARLSLVACAAAPAE-----APLSHVQALAEALAKPKDDPTKHYCVAAVAVY 97  
 DB 11 AGARHVAVAAGPTSECCCPPTPOAKLPHMQALDELAKPKE--SRRLTAQIAVAVRV 68  
 QY 98 AIAETGLGATTPPOLAEGRLRGFEDEVFTLFGADLTMEGSELLHRLTEHLEAHF 157  
 DB 69 AIAETGLGAPDVTIQGLVGLRMLGFDVFTLFRADLTMEGSELLHRLDHLQHP 128  
 QY 158 HSDEPLPMFTSCCPGMIAMLEKSYDPLIFVSSCKSPOMLAAVKSYLEAKGIAPKDM 217  
 DB 129 NKEEPLPMFTSCCPGVAVAVKSNPELLIFLSSCKSPOMLGAIVIKNYAQQVQPSDI 188  
 QY 218 VAVSIMPCRKQSEADRDCEVDADPTLQDLHVTITVELGNIFKRGINLAELPGEMD 277  
 DB 189 CNVSVAPCVKQGEADREWFNTAGLADVDHVTAAEKGITFLRGIKLNLPEPSND 248  
 QY 278 NPMGVSGAGVLEGTGVMERALRAYELFTGTPRLSLSEVRGMDGIKETNITMVA 337  
 DB 249 NIEGSETGALLEGTGGVMEALRTVIEVYVQKPMGRVDFEVRLEEGIKELITLKR 308  
 QY 338 PQSKPELLKHRAAARAAAHGTPGLAMDAGFTSEDRGKITLRVAANGIGNAKK 397  
 DB 309 DDSPEK-----AFAGADGQ-GITLKIAVANGIGNAKK 339  
 QY 398 LITKMOAGEAKYDFEIMACPGCVSGGQPSRDKATQKRAALYNDEKTLKRSHE 457  
 DB 340 LTKSISEGAKTDFEIMACPGGCTGGGQPSRDKQILQKROAMYNDEKSTIRSHD 399  
 QY 458 NPSIRELYDTYLGEPLGKRAHLELHTHYVAGVEEK 493  
 DB 400 NPIQLALYDKPLGAPRSHKADHLHTHYVAGGIPEE 435

RESULT 4  
 Q9AR66 PRELIMINARY; PRT; 449 AA.

ID 09AR66  
 AC 09AR66  
 DT 01-JUN-2001 (TEMBLrel. 17, Created)  
 DT 01-JUN-2001 (TEMBLrel. 17, Last sequence update)  
 DT 01-DEC-2001 (TEMBLrel. 19, Last annotation update)  
 DE Fe-hydrogenase precursor.  
 GN HYDA.  
 OS Scenedesmus obliquus.  
 OC Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Chlorococcales;  
 OC Scenedesmeaceae; Scenedesmus.  
 OX NCBI\_TaxID=3088;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE-21125815; PubMed-11096090;  
 RA Florin L., Tsokogiou A., Happe T.;  
 RT "A novel type of iron hydrogenase in the green alga Scenedesmus obliquus is linked to the photosynthetic electron transport chain."  
 RL J. Biol. Chem. 276:6125-6132(2001).  
 DR EMBL: AJ271546; CAC34419.1; -.  
 DR HSSP: P29166; 1FEH.  
 DR InterPro: IPR004108; Fe\_hyd\_1g\_C.  
 DR InterPro: IPR003149; Fe\_hyd\_SSU.  
 DR Pfam: PF02906; Fe\_hyd\_1g\_C; 1.  
 DR Pfam: PF02256; Fe\_hyd\_SSU; 1.  
 KW Transit peptide.  
 FT TRANSIT 1 35 POTENTIAL.  
 SQ SEQUENCE 449 AA; 48540 MW; 9ABFC5E69580B458 CRC64;

Query Match 53.7%; Score 1384.5; DB 10; Length 449;

Best Local Similarity 58.2%; Pred. No. 1.4e-90;  
 Matches 285; Conservative 52; Mismatches 98; Indels 55; Gaps 9;

QY 11 AVSIRSSCARQVAPAPLAAS---TYAVALATEAPARRIGNVACAAAPAPAPLS 66  
 DB 11 AVSVR-----PVRNRAVVAADERRRLVVAAGPTAECD-----CPPAPAPAP-- 53  
 QY 67 HVOQALAEALAKPKDDPTKHYCVAAVAVVAIAETGLGAPATTPKQLAEGRRGFDE 126  
 DB 54 HMOQTDELAKPKEQ--RKVMAQIAVAVVAIAETGLGAPGVTVGQVYGLRMLGFDY 111  
 QY 127 VFDTLFGADLTMEGSELLHRLTEHLEAHPSDEPLPMFTSCCPGMIAMLEKSYDPL 186  
 DB 112 VFDTLFGADLTMEGSELLHRLDHLQHPNKEEPLPMFTSCCPGVAVAVKSNPELL 171  
 QY 187 YVSSCKSPOMLAAVKSYLEAKGIAPKDMVAVSIMPCRKQSEADRDCEVDADPTLR 246  
 DB 172 YLSSCKSPOMLGAIVIKNYFAAEAGAKPEDICVAVPCVRKQGEADREWFNTGAGG-A 230  
 QY 247 QLDHVTITVELGNIFKRGINLAELPGEKDNPMGVSGAGVLEGTGVMERALRAYE 306  
 DB 231 NVDHVMTTELKIFERGIKINDLOESPDPNPGVSGGGLGTGGVMEALRTVYE 290  
 QY 307 LFTGTPRLSLSEVRGMDGIKETNITMVAAPSKFEELKHRAAARAAAHGTPGPIA 366  
 DB 291 VVTRKPLDRIVFEDVGLGEEKSTLHLPGRPTSP-----KAFAGADG----- 335  
 QY 367 WDGGAGFTSEDRGKITLRVAANGIGNAKKLTQKMOAGEAKYDFEIMACPGCVGGG 426  
 DB 336 -----GITLNIIVANGIGNAKKLTQKMOAGEAKYDFEIMACPGCIGGG 381  
 QY 427 QPRSTKATQKRAALYNDEKTLRSHENPSIETDFTYLGEPLGKRAHLELHTHY 486  
 DB 382 QPRSADKQILQKROAMYNDEKTLRSHENPLIGALAEKFLGEPGKRAHLELHTHY 441  
 QY 487 AGGVEEKDEX 496  
 DB 442 AGGV--PDEK 449

RESULT 5  
 Q9AU60

ID 09AU60 PRELIMINARY; PRT; 403 AA.

ID 09AU60  
 AC 09AU60  
 DT 01-JUN-2001 (TEMBLrel. 17, Created)  
 DT 01-JUN-2001 (TEMBLrel. 17, Last sequence update)  
 DT 01-DEC-2001 (TEMBLrel. 19, Last annotation update)  
 DE Fe-hydrogenase (Fragment).  
 OS Scenedesmus obliquus.  
 OC Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Chlorococcales;  
 OC Scenedesmeaceae; Scenedesmus.  
 OX NCBI\_TaxID=3088;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX SYRAIN-WILD TYPE D3;  
 RC MEDLINE-21292711; PubMed-11400057;  
 RA Wunschlert R., Stangier K., Senger H., Schulz R.;  
 RT "Molecular evidence for a Fe-hydrogenase in the green alga Scenedesmus obliquus."  
 RL Curr. Microbiol. 42:353-360(2001).  
 DR EMBL: AF276706; AAG59621.1; -.  
 DR HSSP: P29166; 1FEH.  
 DR InterPro: IPR004108; Fe\_hyd\_1g\_C.  
 DR InterPro: IPR003149; Fe\_hyd\_SSU.  
 DR Pfam: PF02906; Fe\_hyd\_1g\_C; 1.  
 DR Pfam: PF02256; Fe\_hyd\_SSU; 1.  
 FT NON\_TER 1  
 SQ SEQUENCE 403 AA; 43643 MW; 46110F53C137DA7F CRC64;

Query Match 46.4%; Score 1196; DB 10; Length 403;  
 Best Local Similarity 57.7%; Pred. No. 3.3e-77;  
 Matches 248; Conservative 48; Mismatches 96; Indels 38; Gaps 9;

QY	67	HYQVQALAEALAKKDDPRTKHCVCQVAPAVRAVIAEELGLAPATGTTTKQALAEGRIRGDFDE	126
Db	2	HHQOVLDELAKKE---KRVMLAQIAPAVR-GIAETMGINPGDVTVGQWNTGIRMLGFDY	57
QY	127	VFDTLFGADLTIMEGSELHNRLTETHELAHPHSDEPLPMTSCCPGIMALEKSYFDLIP	186
Db	58	VFDTLFGADLTIMEGTELLHRLDHLQHNPKKEEPLPMTSCCPGVANVEKSNELIP	117
QY	187	YVSSCKSPQMLAAVYKSLAEKKGITAPRDVAVVSIIMPCTRKQSEADRDWFCVDADPTLR	246
Db	118	YLSCKSPQMLGAIVIKNFPAEAGAKRPEDICNVSPVCRKSGEAPPSGSTRHNRGR	177
QY	247	QLDHITVYELNFIKERCINLAELPEGEWDPMGVSGAGVLPFTTGGYMEALRTAYE	306
Db	178	DVDHMTTAEELGKIVYERKIKLNEIQESFDPNVGSGGGGLPFTTGGYMEALRTVYE	237
QY	307	LEFTGPL--PRLSLEVRGMDIKETNITMVAPSGKFEELLKRRAAABAALAAHSTPG-	363
Db	238	VVTAEALPQSRSLT-----TSTAMTPA-----GRASPRSPQAPPAAPSR	277
QY	364	PLANDGGAGFTSEDDRGCTTLTAVVAVANGIGNAKKLLTKAQAGEAKYDEVIACAPAGVG	423
Db	278	PL-----OAOETSGEITTLNIAVANGIGNAKKLLTKQLAAGSKDYDFEVMAKPCGCGIG	328
QY	424	GGGPRSTDKATITQKQALVLYLDEKSTLRSHENPSTIELDYTYGEPFLGHANHELHT	483
Db	329	GGGQPG-RNKQLQKROAAHYDLDERAVYRRT-ENPLIALYELKFLGEPNGHANHELHT	386
QY	484	HYVAGVEEK 493	
Db	387	HYVAGVDPDR 396	
RESULT 6			
ID	Q9WY44	PRELIMINARY; PRT: 608 AA.	
AC	Q9WY44:		
DT	01-NOV-1999 (TREMBLrel. 12, Created)		
DT	01-NOV-1999 (TREMBLrel. 12, Last sequence update)		
DT	01-MAR-2002 (TREMBLrel. 20, Last annotation update)		
DE	MDP-reducing hydrolgenase, subunit D, putative.		
GN	TM0201.		
OS	Thermotoga maritima.		
OC	Bacteria; Thermotogae; Thermotogales; Thermotogaceae; Thermotoga.		
OX	NCBI_Taxid=2336;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=MSB8 / DSM 3109;		
KX	MEDLINE=99287316; PubMed=10360571;		
RA	Nelson K.E., Clayton R.A., Gill S.R., Gwin M.L., Dodson R.J.,		
RA	Hatt D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,		
RA	McDonald L., Utterback T.R., Malek J.A., Linher K.D., Garrett M.M.,		
RA	Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D.,		
RA	Heidelberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,		
RA	Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.,		
RT	"Evidence for lateral gene transfer between Archaea and Bacteria from		
RT	genome sequence for Thermotoga maritima."		
RL	Nature 399:323-329(1999).		
EMBL	AE001705; AAD35293.1; -		
DR	HSSP; P29166; IFEH.		
DR	TIGR; TM0201; -		
DR	InterPro; IPR001450; 4Fe4S_ferredoxin.		
DR	InterPro; IPR001041; Ferredoxin.		
DR	InterPro; IPR001048; Fe_hyd_1g.C.		
DR	InterPro; IPR003149; Fe_hyd_SSU.		
DR	Pfam; PF00111; fer2; 1.		
DR	Pfam; PF00037; fer4; 2.		
DR	Pfam; PF02906; Fe_hyd_1g_C; 1.		
DR	Pfam; PF02256; Fe_hyd_SSU; 1.		
DR	PROSITE; PS00198; 4Fe4S_FERREDOXIN; 1.		
KX	Iron-sulfur, Complete Proteome.		
SO	SEQUENCE 608 AA; 67652 MW; 8C90822E35F905A6 CRC64;		

Query Match	39.2%	Score 1010.5	DB 16	Length 608
Best Local Similarity	47.9%	Pred. No. 1e-63		
Matches 216	Conservative 56	Mismatches 112	Indels 67	Gaps 8
OY	49	IGNVACAAAP--AAEAPLSHVOOALAEALAKPPADDEPTRKHVCQVAPAVAAETLGLAP	107	
DB	218	ICGGGCAACFCPTGALIVENSAYVLELEKE-----KLVVQTAPSRVALGEEFGYAT	272	
OY	108	GATTTCOLAEGRLRGFDPEVPTLGLAOLTTMEESSELLHRLTEHLAHPHSD-EPLPYF	166	
DB	273	GTISGQVAAALRRRGFDVFTNKGADLTTMEESSELERLEK-----GDLEDLPWF	325	
OY	167	TSCCGWIAMLEKSPDILPIYVSSCKSPOMAAVYASIAEKKGIAPRDVWVSIIMPCT	226	
DB	326	TSCCGWVAVLEKYPBELTRLSSAKSPQGMASAVKTYPAKLGKVEDIFHVSIMCT	385	
OY	227	RKQSEADRDWFCVADPTLRQLDHVIITVELGNIFKRGINIAELPEGEWDPNMGVSGA	286	
DB	386	AKKDEALRKQOLVNVGVPV---VDVLLTRLRELKLRMKIKIPANLPEEYDAPLGISTGA	442	
OY	287	GVLPEPTGGVMAALRTIAVELFTGTPRLSLSEVRGMDGKEKNTITWVPAPGSFEBLL	346	
DB	443	AALFVYTGVMALARTIAELTKGALPRLPEEYRGKLVGKAREADIDL-----	490	
OY	347	KHRAAARAAAGTGPRLAMDGAGFTSEDRGGITLRVAVANGAKKILITKQAGE	406	
DB	491	-----DOK---KIRIAVAGSTANVRNMLEILARE	517	
OY	407	AKYDVELIMACAGCGVGGGGRSDKAIYQKROALNLDKSLTRSHSPRSRELYD	466	
DB	518	VKYNHVEVACGCGIGGGOPYSDPILRKRAAATIDERMTRLRSHSPRAIKKLYE	577	
OY	467	TYLSEPLGKHAHELHTHYVAGVHEKEBK 497		
DB	578	EYLEHPLSHKAHELHTY---EDRSRKK 603		
RESULT 7				
ID	Q8RBC8	PRELIMINARY	PRT:	581 AA.
AC	Q8RBC8			
DT	01-JUN-2002	(TREMBlrel, 21, Created)		
DT	01-JUN-2002	(TREMBlrel, 21, Last sequence update)		
DT	01-JUN-2002	(TREMBlrel, 21, Last annotation update)		
DE	NADH dehydrogenase/NADH:ubiquinone oxidoreductase 75 kDa subunit (chain 6).			
GN	NDUG OR TTE0894.			
OS	Thermoaerobacter tengcongensis.			
OC	Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;			
OC	Thermoaerobacteriales; Thermoaerobacteriaceae; Thermoaerobacter.			
NCBI	NCBI_TaxID=119072;			
RU	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN-MB4T / JCM11007;			
RA	MEDLINE=21992816; PubMed=11997336;			
RA	Bao Q., Tian Y., Li W., Xu Z., Xuan Z., Hu S., Dong W., Yang J.,			
RA	Chen Y., Xue Y., Xu Y., Lai X., Huang L., Dong X., Ma Y., Ling L.,			
RT	Tan H., Chen R., Wang J., Yu J., Yang H.;			
RL	"A complete sequence of T. tengcongensis genome."			
DR	Genome Res. 12:689-700(2002).			
EMBL	AE013056; AAM24150.1; -			
Ubl	ubiquinone; Complete proteome.			
KQ	SEQUENCE 581 AA; 64466 MW; BC95C3203A9B1EB7 CRC64;			
Query Match	36.3%	Score 935.5	DB 16	Length 581
Best Local Similarity	43.2%	Pred. No. 2.1e-58		
Matches 189	Conservative 71	Mismatches 110	Indels 67	Gaps 7
OY	52	VACAAAPAAEAPLSHVOOALAEALAKPPDDPTRKHVCQVAPAVAAETLGLAPGTT	111	
DB	199	AVCPYGAATYEDHTRKRYEALA-----DD--KRTVAAQTPAAVAAVAGEEFGMGVGTI	250	
OY	112	PKOLAEGLRRIGFDEVEDTTEGADLTTMEESSELLHRLTEHLAHPHSDPELPYFTSCPP	171	

```

Db      251  TGNMAALRRMGFAVFDTNFAADLTMEEGSELLERI-----KHGKG-LPMITSCSP 302
OY      172  GWIMALEKSTPDLPIPVYSSCKSPQMMLAAMKSYIAEKKGIAPKDMVAVSIMPCTRKROSE 231
Db      303  GWIMALEKSTPDLPIPVYSSCKSPQMMLAAMKSYIAEKKGIAPKDMVAVSIMPCTRKROSE 362
OY      232  ADROMFCVADPTLRQLDHTVITTVYELGNIKFERGINLAELPEGENDNPMVGSAGAVLEFG 291
Db      363  IEREMITNG---KMDVAVLITRLRLAMIKEMGIDFVNLKDEDFELGNTGTGALIFG 419
OY      292  TTGGVMEALRTAVLELFTGTPPLRLSLSEVKGMDIKETNITMVPAPSGKEFELLKHRAA 351
Db      420  ATGCVMEALRTAVLEIYVGRDGIKIDFEVGRGLEGEVREATTITI-----462
OY      352  ARAEAAAGTGPPLAMDAGAGFTSBDGCGITLRLVAVANGIGNAKLITKMOAGRAKYDF 411
Db      463  -----DGMIDKIAIANGTGNMAKLLDKYVAGEVEYHF 494
OY      412  VEIMACPAGCVGGGCGP---RSTDKAITOKROALYNLDEKSTLRSHENPSIRELYDPT 468
Db      495  IEVWCGPGCGIMGGGQPIHNPMEVEVKRAKAIYEIDKNLPIKSHENPAIKRLYEERF 554
OY      469  LGEPLGHAHELILATHY 485
Db      555  LGYPLSEKSHELLATHY 571

```

## RESULT 8

```

O9XC55  PRELIMINARY: PRT: 579 AA.
ID      09XC55
AC      01-NOV-1999 (TREMBLrel. 12, Created)
DT      01-NOV-1999 (TREMBLrel. 12, last sequence update)
DT      01-DEC-2001 (TREMBLrel. 19, last annotation update)
DE      Hydorgenase-1 (EC 1.18.99.1) (fragment).
GN      HYDA.
OS      Clostridium thermocellum.
OC      Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;
OC      Clostridiales; Clostridiaceae; Clostridium.
OX      NCBI_Taxid=1515;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN-ATCC 27405;
RA      Desai S.G., Stevens D.R., Stevenson D.M., Prince H.L., Guerino M.L.,
RA      Lynn L.H.;
RT      "Direct Submission.";
RT      Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
DR      EMBL; AF148212; AAD3071.1; -.
DR      HSP: P00195; 1CLE.
DR      InterPro: IPR001450; 4Fe4S_ferredoxin.
DR      InterPro: IPR001041; Ferredoxin.
DR      InterPro: IPR004108; Fe_hyd_19_C.
DR      InterPro: IPR003149; Fe_hyd_SSO.
DR      Pfam: PF00111; fer2; 1.
DR      Pfam: PF00037; fer2; 2.
DR      Pfam: PF02906; Fe_hyd_19_C; 1.
DR      Pfam: PF02256; Fe_hyd_SSO; 1.
DR      PRINTS: PR00353; 4Fe4S_FERREDOXIN.
DR      PROSITE: PS00198; 4Fe4S_FERREDOXIN; 1.
KW      Iron-sulfur; Oxidoreductase.
FT      NON_TER 579
FT      SEQUENCE 579 AA; 63584 MW; 4D5A5ED4FA526DC7 CRC64;

```

## Query Match

Best Local Similarity 41.1%; Score 903.5; DB 2; Length 579;  
Matches 198; Conservative 61; Mismatches 146; Indels 77; Gaps 7;

```

OY      17  SSCRAQVAPRAPPLASTYVALATL-----EAPARRLGNV--ACAAAPAAEADPLSHY 68
Db      154  SACKNVQTVGALIDYERGRFTVYTAFNKPLSEVPCVNGGCIINCPVAGALREKDDIDYV 213
OY      69  QOALAEIAPKDDPTFKHVQVAVPAVVAIAETIGLAGATTPKQLAGLRLRGDFEYF 128

```

```

Db      214  WEALAN-----PELHVVQTAFAVRAALGEEFGMPIGSRVTKMAALSRLGFKKYF 265
OY      129  DTLFCADLTMEEGSELLRLTEHLLEAPHSDEPLPMTSCCPGMIAMLEKSYPOLIYV 188
Db      266  DDTFADLTMEEGSELLRLTEHLLEAPHSDEPLPMTSCCPGMIAMLEKSYPOLIYV 317
OY      189  SSCRAQVAPRAPPLASTYVALATL-----EAPARRLGNV--ACAAAPAAEADPLSHY 248
Db      318  SSCRAQVAPRAPPLASTYVALATL-----EAPARRLGNV--ACAAAPAAEADPLSHY 374
OY      249  DHVITTVELGNIKFERGINLAELPEGENDNPMVGSAGAVLEFGTTGCVMEALRTAVLEL 308
Db      375  DVVLTTRRLARLIKETGIDFNSLPDKQFDDPGAGSAGAVIFGATGCVMEALRTAVLEL 434
OY      309  TGTPLRLSLSEVKGMDIKETNITMVPAPSGKEFELLKHRAAARAAEAHAHTGPPLAMD 368
Db      435  SGRPADKILEYTVRGLDGIRKASIEL-----460
OY      369  GGAGFTSBDGCGITLRLVAVANGIGNAKLITKMOAGRAKYDFVEIMACPAGCVGGGCGP 428
Db      461  -----DGFLLKAANVANGIGNAKLITKMOAGRAKYDFVEIMACPAGCVGGGCGP 509
OY      429  RSTDKAITOKROALYNLDEKSTLRSHENPSIRELYDPTLGEPLGHAHELILATHY 483
Db      510  IOPSSVNRMKDIRCEAKAIYEDESLPIRKSHENPKIMLYEEFFGEPGSHAKHELILATHY 569
OY      484  HY 485
Db      570  HY 571

```

## RESULT 9

```

O8TG63  PRELIMINARY: PRT: 555 AA.
ID      O8TG63
AC      01-JUN-2002 (TREMBLrel. 21, Created)
DT      01-JUN-2002 (TREMBLrel. 21, last sequence update)
DT      01-JUN-2002 (TREMBLrel. 21, last annotation update)
DE      [Fe]-hydrogenase (fragment).
DE      Hydrogenase sp. E2.
OS      Eukaryota; Fungi; Chytridiomycota; Neocallimastix;
OC      Neocallimastixaceae; Piromyces.
OX      NCBI_Taxid=73868;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN-E2;
RA      MEDLINE-2188625; PubMed-11891051;
RA      Voncken F.G.J., Boxma B., van Hoek A.H.A.M., Akhmanova A.S.,
RA      Vogels G.D., Huynen M., Veenhuis M., Hackstein J.H.P.;
RT      "A hydrogenosomal [Fe]-hydrogenase from the anaerobic chytrid
RT      Neocallimastix sp. L2."
RL      Gene 284:103-112(2002).
DR      EMBL; AF446076; AAL90459.1; -.
FT      NON_TER 1
FT      SEQUENCE 555 AA; 62033 MW; AEF0134FBB863B18 CRC64;

```

Query Match 34.7%; Score 893; DB 3; Length 555;  
Best Local Similarity 44.5%; Pred. No. 2,1e-55;  
Matches 181; Conservative 67; Mismatches 105; Indels 54; Gaps 7;

```

OY      81  DPTFKHVQVAVPAVVAIAETIGLAGATTPKQLAGLRLRGDFEYFDTLFCADLTME 140
Db      186  DTRKRTVVSTAPATVATVAALEEFNAPDPDTGKMAVAGAKLGFDTLFDNRSADLTME 245
OY      141  ESESELRLRLTEHLLEAPHSDEPLPMTSCCPGMIAMLEKSYPOLIYVSSCKSPQMMLA 200
Db      246  EGTLELTRLE-----GCKEPMTSCCPGMIAMLEKSYPOLIYVSSCKSPQMMLA 297
OY      201  MVSYSIAEKKGIAPKDMVAVSIMPCTRKROSEADRMFCVADPTLRQLDHTVITTVELGNI 260
Db      298  VIKTYPARKINAKPEDIIHVSVPCTAKKGAEKRPFKRDYV---DIDHVITTRRLITL 354

```

```

OY 261 FKRGINLAELPEGENDNPGVSGAGVLGTTGVMERALRTAYELFTG--PTLRRLSL 318
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 355 LKLRINPELKEKEDSPGLGSSAGNLEFGVTGVMERALRTA-QITGVENPILDEL 413
OY 319 SEVRGMDGKIKETNITNVPAPGSKFEELKRRARAALAAHGTPLANDGAGFTSEDO 378
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 414 KAIRGIDGIRKASV-----PL-----KTKDG 434
OY 379 RRGITLRVAVANGNAGKLLITKMOAGEAKYDFEIMACPGAGGCGPRSDKATLTK 438
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 435 K-DVNRRAAVVSGANIQFLKLEKLEPFDEVMKCGGCGINGGCPKSDAPRYAK 493
OY 439 ROALYNLDEKSTLRSHENPSIRELYDTYLGEPILGHAKHELHHTHY 485
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 494 KMERMTMDQASLRSLSHENPSITQYKEPLKEPNHLSHELLHHTHY 540

```

```

RESULT 10
OY 08TFP2 PRELIMINARY; PRT; 636 AA.
AC 08TFP2:
DT 01-JUN-2002 (TEMBLrel. 21, Created)
DT 01-JUN-2002 (TEMBLrel. 21, Last sequence update)
DE 01-JUN-2002 (TEMBLrel. 21, Last annotation update)
GN Hydorgenase.
OS Neocallimastix frontalis (Rumen fungus).
OC Eukaryota; Fungi; Chytridiomycota; Neocallimastixales;
OC Neocallimastixaceae; Neocallimastix.
OX NCBI_TaxID=4757;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21888625; PubMed=11891051;
RA Voncken F.G.J., Boxma B., Van Hoek A.H.A.M., Akhmanova A.S.,
RT Vogels G.D., Huynen M., Veenhuis M., Hackstein J.H.P.,
RT "A hydrogensomal [Fe]-hydrogenase from the anaerobic chytrid
RL Neocallimastix sp. L2."
DR EMBL; AY033895; AAK60409.1;
SQ SEQUENCE 636 AA; 70449 MW; 4A4E78235FAA0D7 CRC64;

```

```

Query Match 34.0%; Score 880.5; DB 3; Length 636;
Best Local Similarity 43.0%; Pred. No. 2e-54;
Matches 168; Conservative 67; Mismatches 123; Indels 59; Gaps 9;
OY 54 CAAAP-AAEAPLSHYQALAEIAELAKPKDPTRKHVCVQVAPAVVAIAETLGLAPGATTP 112
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 238 CQGVCPVGAITATEVVDYLRHL-----DTKRKYVVCSTAPAIRVAPAEFSEADPDT 292
OY 113 KOLAEGLRLGDEVEFDTLLEGADLTMEGSELLRLTEHLNAPHSDEPLPFTSCCPG 172
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 293 GKXVVALRKIGFYIDTNSADLTMEGTELDRL-----NNGCKPFTSCCPG 344
OY 173 WTAMLEKSYPDILPYVSSCKSPOMLAAMVKSYLAKKGIAPKDMVMSIMPTKQSEA 232
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 345 WTNWKSSTPELSDNLSCKSPOMIGAVIKSTFAKLGISTEDIIIVSIMPTAKGSA 404
OY 233 DRDMFCVDA--DPTLRQLDHVITVYELGNIIFKRGINLAELPEGENDNPGVSGAGVLF 290
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 405 RREFEYQKRGKDDYDIDYITTRRLTLKLLKAKINPAELPDDKDSPLGISTSSGNLF 464
OY 291 GTTGCVMAALRTAYELFTG--TPLRSLSEVRGMDGKIKETNITNVPAPGSKFEELKH 348
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 465 GYVGCGVMAAIRTA-QVITGVENPILGLKAIKRGIDGIRKASV-----507
OY 349 RAAARAAAHGHPGLANDGAGFTSEDRGKITLRVAVANGNAGKLLITKMOAGEAK 408
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 508 -----PL-----KTKDG--EVSRAAVVSGAGAIQFLKIRKKELE 543
OY 409 YDFEIMACPGAGGCGPRSDKATLTKMOALNLDKSTLRSHENPSIRELYDTY 468
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 544 PFIEEMKCGGCGINGGCPKSDAPRYAKKMRMTMDQAKRLCHENPELIDYKKNF 603

```

```

OY 469 LGEPILGHAKHELHHTHY 485
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 604 LGEPNSHLAHELHHTHY 620

```

```

RESULT 11
OY 046508 PRELIMINARY; PRT; 585 AA.
AC 046508:
DT 01-NOV-1996 (TEMBLrel. 01, Created)
DT 01-NOV-1996 (TEMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TEMBLrel. 19, Last annotation update)
DE Potential NAD-reducing hydrogenase subunit.
OS Desulfovibrio fructosovorans.
OC Bacteria; Proteobacteria; delta subdivision; Desulfovibrionaceae;
OC Desulfovibrio.
OX NCBI_TaxID=878;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95270577; PubMed=7751270;
RA Malki S., Salimaine I., De Luca G., Rousset M., Derroun Z.,
RA Belatch U.P.;
RT "Characterization of an operon encoding an NADP-reducing hydrogenase
RT in Desulfovibrio fructosovorans."
RL J. Bacteriol. 177:2628-2636(1995).
DR EMBL; U07229; AAA87057.1;
DR HSSP; P29166; 1FEH.
DR Interpro: IPR001450; 4Fe4s_ferredoxin.
DR Interpro: IPR001041; Ferredoxin.
DR Interpro: IPR004108; Fe_hyd_19_C.
DR Interpro: IPR003149; Fe_hyd_SSU.
DR Pfam: PF00111; fer2; 1.
DR Pfam: PF00037; fer4; 1.
DR Pfam: PF02906; Fe_hyd_19_C; 1.
DR Pfam: PF02286; Fe_hyd_SSU; 1.
DR PRINTS; PRO0333; 4FE4SFERDOXIN.
DR PROSITE; PS00198; 4FE4S_FERREDOXIN; 1.
KW Iron-sulfur.
SQ SEQUENCE 585 AA; 63429 MW; C14D0E6A3E2A72 CRC64;

```

```

Query Match 34.0%; Score 877; DB 2; Length 585;
Best Local Similarity 41.4%; Pred. No. 3.1e-54;
Matches 192; Conservative 66; Mismatches 138; Indels 68; Gaps 9;
OY 33 STVRVALATLEAPARRLGNYA-----CAAAP-AAEAPLSHYQALAEIAELAKPKDP 82
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 173 SCVNRGFTAVNAPAFEM-NADRVTCNCGCVAVCPGALVEHEIYMEVEALAND--- 228
OY 83 TRKHVCVQVAPAVVAIAETLGLAPGATTPKOLAESLRRLGDEVEFDTLLEGADLTMEG 142
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 229 --KVIVVQTPAARVALGIEDLVAPGTSVTKMAALRLRGPFHDVDTQDAADLTMEEG 286
OY 143 SELHLRLTEHLNAPHSDEPLPFTSCCPMIMLEKSPDILPYVSSCKSPOMLAAMV 202
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 287 SEFLDLRLGHLAC--DTNKKPLTLTSCCPWVFFEHQFDMADVSTASPOOMGALA 344
OY 203 KSYLAEEKGIAPKDMVMSIMPTKQSEADRDMPCVDA-DPTLRQLDHVITVYELGNI 262
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 345 KTYVALDLGIPRKLVVSVMPCLAKKYECAREFSVNGNP--DVDIYITTELAKLVK 401
OY 263 ERGINLAELPEGENDNPGVSGAGVLFCTTGVMERALRTAYELFTGTPRLPLSSEVR 322
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 402 RNIIDFAGLPDEEDFPLGASTGAAPIFGVGTGVIEMALRTAYELATGETLAKVEDEV 461
OY 323 GMDGKIKETNITNVPAPGSKFEELKRRARAALAAHGTPLAMGCGAGFTSEDRGGI 382
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 462 GMDGVAKKAKVY-----GDN 476
OY 383 TLRVAVANGNAGKLLITKMOAGEAKYDFEIMACPGAGGCGGP-RSTDKAITOKRA 441
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 477 ELVIGVAHNGNRELKPCGAGET--FNAIEVACPGGCGGCGGYRHGDVVELKKRTO 535
OY 442 ALYNLDEKSTLRSHENPSIRELYDTYLGEPILGHAKHELHHTHY 485

```



<b>RESULT 14</b>					
ID	O9GTP2	PRELIMINARY:	PRT:	589 AA.	
AC	O9GTP2:				
DT	01-MAR-2001 (TREMBLrel. 16, Created)				
DT	01-MAR-2001 (TREMBLrel. 16, Last sequence update)				
DT	01-JUN-2002 (TREMBLrel. 21, last annotation update)				
DE	Purative 64kda Iron hydrogenase (Fragment).				
OS	Trichomonas vaginalis.				
OC	Eukaryota; Parabasalidae; Trichomonadida; Trichomonadidae;				
OC	Trichomonas.				
OX	NCBI_TaxID=5722;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE=20523972; Pubmed=11070057;				
RA	Horton D.S., Foster P.G., Embley M.T.;				
RT	"Iron Hydrogenases and the Evolution of Anaerobic Eukaryotes.";				
RL	Mol. Biol. Evol. 17:1695-1709(2000).				
DR	EMBL: AF262401; AAC31037.1; -				
DR	HSSP: P29166; IFRH				
DR	InterPro: IPR001450; 4Fe4S_ferredoxin.				
DR	InterPro: IPR001041; Ferredoxin.				
DR	InterPro: IPR004108; Fe_hyd_19_C.				
DR	InterPro: IPR003149; Fe_hyd_SSU.				
DR	pfam: PF00111; fer2; 1.				
DR	pfam: PF00037; fer4; 2.				
DR	pfam: PF02906; Fe_hyd_18_C; 1.				
DR	pfam: PF02256; Fe_hyd_SSU; 1.				
DR	PROSITE: PS00198; 4FE4S_FERREDOXIN; 1.				
FT	NON_TER				
SQ	SEQUENCE 589 AA; 64854 MW; F0B69ED78B9D066A CRC64;				
Query Match	31.3%; Score 805.5; DB 5; Length 589;				
Best Local Similarity	36.2%; Pied. No. 4e-4;				
Matches 177; Conservative 79; Mismatches 132; Indels 101; Gaps 8;					
OY	6 LKPCAASVIRSGSCCARQAVAPAPAPLAASIVRVALATLEAPARRLDGNVCAAAAPAAREPL 65				
DB	187 IQPTFGVTLQETSC-----IKGGCTLYCPV-----GAITEK 218				
OY	66 SHVOQALAEIAKKPKDDPTRRKHVCYOVAVPAVRVAIAETGLIAPGATTPKOLAEGRRLGFD 125				
DB	219 SQVKRALDIILAE----NKGKITTVQVAAVPAVRVAISAEAGCYKGYTTTGMVSALAALGFD 274				
OY	126 EVFDTFLFGADLTIMEEGSELLRHLETHEAPHSDPELPMTTSCPCGWIAMLEKSPDLI 185				
DB	275 LYVDITNGADLLICEAGELVRLRD-----PAKFPMFTTCPCAWNVNYEQSAPDRI 327				
OY	186 FYVSSCKSPOMLAAMVKSSTLAEEKGIAPKDMWVNSIMPCTRKKGSADRDHFCDVADPLL 245				
DB	328 PNLSSCRSPQCMLSALKINYPLKLIDVQOEVLNFSIMPCTAKKEVER-----DEL 379				
OY	246 R-----QLDHVTITVELGINFKERGINMAELPEBEENPNMGVSGAGVLFCTTGWMEA 299				
DB	380 RTKSLKLTWDVLTVALRELVEMKISINDENNLPDPQDNIFEGSGAGQIFAATGVMEA 439				
OY	300 ALRTAYELTEGTPLRLSLSEVRGADGJKETRTITVPAWGSKFEELLKRHAARAABAANAH 359				
DB	440 ASRTAFEVYTGTGKLNNVINIYPKRGDGLRIALLDL----- 474				
OY	360 GTPGCLANDGAGGFTSEDGRCGITLRVAVANGCLGNAKKLITRMQAQEA--YDEVETWA 416				

Query Match	30.8%; Score 794.5; DB 2; Length 606;
Best Local Similarity	42.4%; Pred. No. 2.5e-48;
Matches	169; Conservative 61; Mismatches 112; Indels 57; Gaps 6
QY	90 QVAPAVVAIAETTLGLACGATTPKOLABGLRGLGFEVFTLFGADLTIMEGSELLHRL 149
DB	233 QFAPAVRVLGEGFGLPPGSSVEGVPTALRLIGADVLDITNFADLVIMEGETELLORTL 292
QY	150 TEHLEAHHSDEPLDMFTSSCCGWTAMLEKSTPDLIPYSSCKSSQPMMLAAVYKSLAK 209
DB	293 -----RGCAKLTLFTLFTSCCGWTFEKKHLPDILPHVSTTRSPQCCGALAKTYLAT 344
QY	210 KGIAKDWVMSIMFCTRKQSEADRDWCVDADPTLRQLDHYITVTEGNIKEKRGINTA 269
DB	345 MNVAERRRNVVSLMCTAKKEEAARPERFRDS---VRDDVALTTREARLLRREGCIDIA 401
QY	270 ELPEGEMPN--MGVSGGAGVLTEGTGVEALRLTAYELFTGTPLPRLSLSEVRGMDGIR 328
DB	402 GLESPGCDPLMGKRAITGAAVITGTCGVMEALRLVYHNLNKELAPVLEHLMLRGYENR 461
QY	329 ETNITMVPAPGSKFEELLKHPRAAARAAAAGCTPGPLAWDGAGFTSDEGCGITLRVAV 388

		1. :		11		1 :	1111	
Db		462	EAVV----	PL-----			GEENGSKVAV	478
Oy		389	ANGLSNAKLLITKMOAGEAKYDFVEIMACPACGVGGGPRS-	TDKAITOKRQAALYNL				446
			11 : 111 :	11 : 11111111 :	11111111 :			
Db		479	VHGKKAQOMEAVYLAGRADHFVEYVMACPGCGMGGGGPRSKRAYNPNAQARRAALFSL					538
Oy		447	DEKSLRSHENPSIRELYDFYLGEPILGKAHEILHTHY					485
			11 : 11111 :	11 : 11111111 :	11 : 11111 :			
Db		539	DAENALROSHNNPLIGKVFESFLGEPSCSLSHRLHTRY					577

```
Search completed: June 3, 2003, 16:09:07
Job time : 88 secs
```

**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 3, 2003, 15:38:18 ; Search time 27 Seconds

(without alignments)  
541.600 Million cell updates/sec

Title: CAC80065

Perfect score: 2576

Sequence: 1 MSALYIKPCAAYSIRGSSCR.....HELLHRYVAGVEKDEKK 497

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database :

Issued Patents AA:\*  
1: /cgn2\_6/prodata/1/1aa/5A.COMB.pep:\*  
2: /cgn2\_6/prodata/1/1aa/5A.COMB.pep:\*  
3: /cgn2\_6/prodata/1/1aa/5A.COMB.pep:\*  
4: /cgn2\_6/prodata/1/1aa/5A.COMB.pep:\*  
5: /cgn2\_6/prodata/1/1aa/5A.COMB.pep:\*  
6: /cgn2\_6/prodata/1/1aa/5A.COMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	125.5	4.9	748	4 US-09-066-046-4	Sequence 4, Appli
2	112	4.3	4472	2 US-08-804-227C-2	Sequence 2, Appli
3	109.5	4.3	857	4 US-08-887-534A-47	Sequence 47, Appli
4	105.5	4.1	3481	2 US-07-642-734C-2	Sequence 2, Appli
5	105.5	4.1	3481	2 US-08-439-009A-2	Sequence 2, Appli
6	100.5	3.9	3170	4 US-09-036-987A-4	Sequence 4, Appli
7	100.5	3.9	3170	4 US-09-370-700-4	Sequence 4, Appli
8	99.5	3.9	638	2 US-08-426-125-4	Sequence 4, Appli
9	99.5	3.9	638	2 US-08-455-355-4	Sequence 4, Appli
10	99.5	3.9	3724	2 US-08-804-227C-10	Sequence 10, Appli
11	99.5	3.9	3724	2 US-08-804-198-4	Sequence 4, Appli
12	99.5	3.9	4550	2 US-08-804-227C-8	Sequence 8, Appli
13	99.5	3.9	4550	2 US-08-804-198-2	Sequence 2, Appli
14	98.5	3.8	823	4 US-09-134-001C-4081	Sequence 4081, Ap
15	94.5	3.7	1996	2 US-08-804-227C-9	Sequence 9, Appli
16	94.5	3.7	1996	2 US-08-804-198-3	Sequence 3, Appli
17	94	3.6	537	3 US-08-920-610-2	Sequence 2, Appli
18	94	3.6	805	4 US-09-513-783A-178	Sequence 178, App
19	94	3.6	2035	1 US-08-046-585-5	Sequence 5, Appli
20	94	3.6	2035	1 US-08-393-703-5	Sequence 5, Appli
21	94	3.6	2035	5 PCT-US93-11721-5	Sequence 5, Appli
22	94	3.6	4545	2 US-08-804-227C-14	Sequence 14, Appli
23	93.5	3.6	3782	4 US-08-105-537-4	Sequence 4, Appli
24	93	3.6	4551	3 US-09-320-878-1	Sequence 1, Appli
25	93	3.6	4613	4 US-09-105-537-31	Sequence 31, Appli
26	93	3.6	11877	4 US-09-105-537-6	Sequence 6, Appli
27	92.5	3.6	514	4 US-09-385-028-5	Sequence 5, Appli

28	92.5	3.6	530	3 US-09-222-817-12	Sequence 12, Appli
29	92.5	3.6	530	3 US-09-222-817-14	Sequence 14, Appli
30	92.5	3.6	530	4 US-09-222-786-12	Sequence 12, Appli
31	92.5	3.6	530	4 US-09-222-786-14	Sequence 14, Appli
32	92.5	3.6	1833	4 US-08-621-944A-4	Sequence 4, Appli
33	92.5	3.6	1833	4 US-08-945-567D-4	Sequence 4, Appli
34	92.5	3.6	1992	4 US-08-621-944A-3	Sequence 3, Appli
35	92.5	3.6	1992	4 US-08-945-567D-3	Sequence 3, Appli
36	92.5	3.6	2048	4 US-09-268-347-48	Sequence 48, Appli
37	92.5	3.6	2123	4 US-08-968-685A-10	Sequence 10, Appli
38	92.5	3.6	3588	4 US-09-036-987A-6	Sequence 6, Appli
39	92.5	3.6	5588	4 US-09-370-700-6	Sequence 6, Appli
40	92	3.6	607	3 US-09-335-409-9	Sequence 9, Appli
41	92	3.6	607	4 US-09-568-102-9	Sequence 9, Appli
42	92	3.6	607	4 US-09-567-969-9	Sequence 9, Appli
43	92	3.6	607	4 US-09-568-480-9	Sequence 9, Appli
44	92	3.6	607	4 US-09-568-486-9	Sequence 9, Appli
45	92	3.6	607	4 US-09-568-472-9	Sequence 9, Appli

#### ALIGNMENTS

RESULT 1  
US-09-066-046-4  
Sequence 4, Application US/09066046A  
Patent No. 6204252  
GENERAL INFORMATION:  
APPLICANT: MURPHY, Cheryl  
STOREY, James  
BELTZ, Gerald A.  
COUGHLIN, Richard T.  
TITLE OF INVENTION: CHARACTERIZATION OF GRANULOCYTIC  
ERLICHIA AND METHODS OF USE  
NUMBER OF SEQUENCES: 65  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: HALE AND DORR LLP  
STREET: 60 State Street  
CITY: Boston  
STATE: MA  
COUNTRY: United States  
ZIP: 02109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION NUMBER: US/09/066,046A  
FILING DATE: 24-Apr-1998  
CLASSIFICATION: <Unknown>  
ATTORNEY/AGENT INFORMATION:  
NAME: Superko, Colleen  
REGISTRATION NUMBER: 39, 850  
REFERENCE/DOCKET NUMBER: 106,941,155  
TELEPHONE: (617) 526-6000  
TELEFAX: (617) 526-5000  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 748 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
SEQUENCE DESCRIPTION: SEQ ID NO: 4:  
US-09-066-046-4  
Query Match 4.9%, Score 125.5; DB 4; Length 748;  
Best Local Similarity 23.2%; Pred. No. 0.00055;  
Matches 66; Conservative 33; Mismatches 104; Indels 81; Gaps 12;

QY 248 LDHVTVEICNIFKRGIMLAELPEGEWNPVCGAGVLEFGTGVMEALRTAYEL 307  
 DB 98 LEHITIT-EVIVSNEE-----ITPBEK-----KTLTAEALTSKGVYKALKNSADV 146  
 QY 308 FTGTPRLSL-----SEVRGMDGKENTITVNPAPGS----- 340  
 DB 147 -NMSPEPATLGLGRCFQSGSKAIKRLRVEAGAHINTPTGSMPLAAVQAAANEASNL 205  
 QY 341 -----KEFELLKRRAAABAAAHGTP-----GPIAWDGGAGFTSEDRGCGIT 383  
 DB 206 KEANKIYNFLHRCADLSTEHCTPALHLATAGNRTAMLLDKGAPATORDAKRRTA 265  
 QY 384 LRAVANGIGNAKKILITKMOAGKAYDVEIMACPGACGGCGPRTSTDAITQKROAL 443  
 DB 266 LHIAANGDGKLYRMIAK-----KCPDSC-----QPLCSMDGDTALHE-AL 305  
 QY 444 Y--NLDEK-----STLRSHENPSTRELYDTLGEFLGKAHELLH 482  
 DB 306 YSDNVEKECFLKMLKESRKLHNSFFGDLNTPQEAANGDTLLH 349

## RESULT 2

US-08-804-227C-2  
 : Sequence 2, Application US/08804227C  
 : Patent No. 5876991  
 : GENERAL INFORMATION:  
 : APPLICANT: Dehoff, Bradley S.  
 : APPLICANT: Kubstoss, Stuart A.  
 : APPLICANT: Rostock, Paul R., Jr.  
 : APPLICANT: Sutton, Kimberly L.  
 : TITLE OF INVENTION: POLYKETIDE SYNTHASE GENES  
 : NUMBER OF SEQUENCES: 15  
 : CORRESPONDENCE ADDRESS:  
 : ADDRESSEE: THOMAS G. PLANT 1501  
 : STREET: LILLY CORPORATE CENTER  
 : CITY: INDIANAPOLIS  
 : STATE: IN  
 : COUNTRY: USA  
 : ZIP: 46285  
 : COMPUTER READABLE FORM:  
 : MEDIUM TYPE: Floppy disk  
 : OPERATING SYSTEM: IBM Compatible  
 : SOFTWARE: ASCII(DOS) Text only  
 : CURRENT APPLICATION DATA:  
 : APPLICATION NUMBER: US/08/804,227C  
 : FILING DATE: February 21, 1997  
 : CLASSIFICATION: 435  
 : ATTORNEY/AGENT INFORMATION:  
 : NAME: Plant, Thomas, G.  
 : REGISTRATION NUMBER: 35,784  
 : REFERENCE/DOCKET NUMBER: X-8231  
 : TELECOMMUNICATION INFORMATION:  
 : TELEPHONE: 317-276-2459  
 : INFORMATION FOR SEQ ID NO: 2:  
 : SEQUENCE CHARACTERISTICS:  
 : LENGTH: 4472 amino acids  
 : TYPE: amino acid  
 : TOPOLOGY: unknown  
 : MOLECULE TYPE: protein  
 : US-08-804-227C-2

Query Match 4.38; Score 112; DB 2; Length 4472;  
 Best Local Similarity 23.48; Pred. No. 0.29;

Matches 91; Conservative 46; Mismatches 124; Indels 128; Gaps 24;

QY 143 SELHRLTEHLEAHP-----HSDPELPMTSCC--PGWIMLEKSYDPLT--- 185  
 DB 1032 AQLRH--TELLDGSFSGSVLAPAKSPFAQEPITAVVGMGRFPGVSPALMLRYVEGV 1089  
 QY 186 -----PY-----VSSCKSPOMMLAANVKSYL-----AEKKIAPKDNVVS 221

DB 1090 DAVSPFQDRGMDVEGLYDEPVGAG--KSYVREGFLHDAEFDAAEFPGISPREAVAMD 1147  
 QY 222 IMPCTRKQSEADRMFCVD---ADPTLRQDHPVITVTELGIFKRGINLAELPEGEWNP 278  
 DB 1148 -----PQORLLLETMEALIERAGIDP---HSLHGSTGYAGVMPQETPRALAEAE----- 1196  
 QY 279 PMGVSGAGVLEFGTGVMEALRTAYELFTGTPLP-----RLSLSEVRG-- 323  
 DB 1197 -----GSDGYLLTGTSGSVSG--RVAVTLGLEPAAVVDTRACSSSLVALHLAVALNGE 1250  
 QY 324 MDGKETNITVNPAPSGKFEELKRRAAA---RAEAAAHGTPGLANDGGAGF----- 373  
 DB 1251 CDNALAGVTVMAQPG--MFVEFSRQGLADGRCARADGADG--TAAAEAGVVLVERLS 1308  
 QY 374 -----TSEDG--RGITPL-----VVAVANGIGNAKKLITKMOAGEAK 408  
 DB 1309 DARRLGHVLAUVGSAVNODGASNGILAPSGSEQERY-IROALGNRLTVADVVEAH 1367  
 QY 409 -----YDFVEIMACPGACVG--GGGCP 428  
 DB 1368 GTGTRLDPIEAQAL-LGTYGRDRDGGRP 1395

## RESULT 3

US-08-887-534A-47  
 : Sequence 47, Application US/08887534A  
 : Patent No. 6455323  
 : GENERAL INFORMATION:  
 : APPLICANT: Holden, David W.  
 : TITLE OF INVENTION: ANTI-BACTERIAL METHODS AND MATERIALS  
 : NUMBER OF SEQUENCES: 106  
 : CORRESPONDENCE ADDRESS:  
 : ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
 : STREET: 233 South Wacker Drive/6300 Sears Tower  
 : CITY: Chicago  
 : STATE: Illinois  
 : COUNTRY: United States of America  
 : ZIP: 60606  
 : COMPUTER READABLE FORM:  
 : MEDIUM TYPE: Floppy disk  
 : OPERATING SYSTEM: IBM PC compatible  
 : SOFTWARE: PC-DOS/MS-DOS  
 : CURRENT APPLICATION DATA:  
 : APPLICATION NUMBER: US/08/887,534A  
 : FILING DATE:  
 : CLASSIFICATION: 536  
 : ATTORNEY/AGENT INFORMATION:  
 : NAME: Rin-Laures, Li-Hsien  
 : REGISTRATION NUMBER: 33,547  
 : REFERENCE/DOCKET NUMBER: 28341/33996  
 : TELECOMMUNICATION INFORMATION:  
 : TELEPHONE: (312) 474-6300  
 : TELEFAX: (312) 474-0448  
 : TELEX: (312) 474-6600  
 : INFORMATION FOR SEQ ID NO: 47:  
 : SEQUENCE CHARACTERISTICS:  
 : LENGTH: 857 amino acids  
 : TYPE: amino acid  
 : TOPOLOGY: linear  
 : MOLECULE TYPE: protein  
 : US-08-887-534A-47

Query Match 4.38; Score 109.5; DB 4; Length 857;  
 Best Local Similarity 21.28; Pred. No. 0.033;

Matches 86; Conservative 67; Mismatches 162; Indels 91; Gaps 19;

QY 108 GATTPKQIACGLRR-LGFDFVFTLFGADLTMEESSELIHRLTEHLEAHPH--SDEPL 163  
 DB 312 GATTLDEYRYIERKDALERFQKVFAEBSV-EDTIALIRGLKERELHHHQTTPAI 370  
 QY 164 PMFTSCPGWIMLEKSYSP---DLIPYSSCKSPON-----MLA 199

Db 371 VAAATLSHRYIA--DROLPDKAIDLIDEAASSIRMOJDSKPEELDRRLDRIITOLKLEQQA 428  
 Qy 200 AMVKSVLAEEKGIADPDWVMSIMPTCKROSEADRDWFCVDAD---PTLR-OLDHVIT 254  
 Db 429 LMKESDEASKRL---DMLNELSDKERYSLELEEMKAKASLSCTOTIKLEQAKIA 485  
 Qy 255 VE---LGNIFKERGINLAELPEGEWDMNMGVSGAVLFGT---GGVMEALRTAY 305  
 Db 486 IQQARVGDGLARMSLELOKIKIPELEKQL-----EATQLEGKTMRLRKRYDALEIAEVL 540  
 Qy 306 ELFTGTPLRLSLSEVRGMDGKEKINITVWPAPGSKFEELKHRAARAEA----- 356  
 Db 541 AMWTGIPVSRMMESE-----REKLRL-----EOELHHRVIGQNEAYDAVSNMIR 585  
 Qy 357 -AAHGTGPGPLANDGAGFTSEDRGITLRVAVANGLSNAKLTITKMOGE--AKYDFVE 413  
 Db 586 NSRAGLADPNRPIGSEFLFGPTGVGKTELCKALANMFSDAMRVIDSEPMERHSVR 645  
 Qy 414 IMACPAGCVG---GGGPRSTDKAITOKRQALYNLDEKSTLRSH 456  
 Db 646 LVGAPPGVGYEGEGYLTEA-----VRRRPYSVILLDE---VEKAK 683  
 Db  
 RESULT 4  
 US-07-642-734C-2  
 ; Sequence 2, Application US/07642734C  
 ; Patent No. 5824513  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Katz, L  
 ; APPLICANT: Donadio, S  
 ; APPLICANT: McAlpine, J B  
 ; TITLE OF INVENTION: Recombinant DNA Method for Producing  
 ; TITLE OF INVENTION: Erythromycin Analogs  
 ; NUMBER OF SEQUENCES: 27  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Edward H. Gorman  
 ; STREET: Abbott Laboratories D377/Ab6D-2 One Abbott  
 ; STREET: Park Rd  
 ; CITY: Abbott Park  
 ; STATE: IL  
 ; COUNTRY: US  
 ; ZIP: 60064-3500  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; OPERATING SYSTEM: IBM PC compatible  
 ; SOFTWARE: Patent Release #1.0, Version #1.25  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/07/642,734C  
 ; FILING DATE: 17-JAN-91  
 ; CLASSIFICATION: 435  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Danckerts, Andreas M  
 ; REGISTRATION NUMBER: 32652  
 ; REFERENCE/DOCKET NUMBER: 4952.US.01  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 708-937-9396  
 ; TELEFAX: 708-938-2623  
 ; INFORMATION FOR SEQ ID NO: 2:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 3491 amino acids  
 ; TYPE: amino acid  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: protein  
 ; US-07-642-734C-2  
 Query Match 4.1%; Score 105.5; DB 2; Length 3491;  
 Best Local Similarity 21.6%; Pred. No. 0.93;  
 Matches 117; Conservative 60; Mismatches 213; Indels 151; Gaps 25;

Qy 66 SHVOAL---AELAKRDPDTRKHVCQVAPAVRAIAETLG----- 104  
 Db 457 --VSAIIGVEVIALIYHPTRPALAEALAGTEVAQRTTRARTENAPGEVAVYAMAC 514  
 Qy 105 -LAPGATPKQ---LAELRL-----GDEVEDTLFGADLTMEGSELLRLTEH 152  
 Db 515 RLPGVSTPEEFELLSEBRDAVAGLPTRGMD--LDSLFHPDPT--RSQT----- 561  
 Qy 153 LEAHPHSDPLPMFTSCCGWIMLEKSTPDILIPVSSCKSQOMLAAMVSKYLAEKKI 212  
 Db 562 --AHQGGGFLTEATFADPAFGM-----SPREALAVDPOQRIMLELSMEVLERAGI 611  
 Qy 213 APKDMVMSIMPTCKROSEADRDWFCVDADPTLRDHLITVTELGINIKERGINIAELP 272  
 Db 612 PPTSL-----QASP-----TGVEGLIPOETGPRLAE--- 638  
 Qy 273 EGEWDMNMGVSGAVLFGTGGVMEALRTAYELFTGTPRLSLSEV----- 321  
 Db 639 -----GGEVGEYLMGTTSV--ASGRIVTL--GLEGPALISVDTCSSSLVAVHLA 687  
 Qy 322 ----RGMDIKET-NITVWPAPGSKFEELKHRAARAEMAAHGTGPGPLANDGAGFTS 375  
 Db 688 COSLRGESSLAMAGGVTVMPGMLYDFSRMKNSLAPDCRCAFSK-----GANGFCM 740  
 Qy 376 EDGRGITLRVAVANGLSNAKLTITKMOGEAKYDFVEI-MACPACVGGGGOPRSTDKA 434  
 Db 741 AEGAGMLLE-RISDRRRNGHPVLAVLRGTAVNSDASNGLSAP-----NGRAQVRITQQA 795  
 Qy 435 ITOK--RQALYNLDEKSTLRSHENPSIRELYDTV---LGEP--LGRHAHELLHTHYVA 487  
 Db 796 LAESGIGRADIOAVEAHGTGTLGDIETEARLFEAVGRBQEPHLIGSVKSNLGHQAAA 855  
 Qy 488 G 488  
 Db 856 G 856  
 Db  
 RESULT 5  
 US-08-439-009A-2  
 ; Sequence 2, Application US/08439009A  
 ; Patent No. 6004787  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Donadio, S  
 ; APPLICANT: Katz, L  
 ; APPLICANT: McAlpine, J B  
 ; TITLE OF INVENTION: Method of Directing Biosynthesis of  
 ; TITLE OF INVENTION: Specific Polyketides  
 ; NUMBER OF SEQUENCES: 27  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Steven F. Weinstock  
 ; STREET: Abbott Laboratories D377/Ab6D-2 One Abbott  
 ; STREET: Park Rd  
 ; CITY: Abbott Park  
 ; STATE: IL  
 ; COUNTRY: US  
 ; ZIP: 60064-3500  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; OPERATING SYSTEM: IBM PC compatible  
 ; SOFTWARE: Patent Release #1.0, Version #1.25  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/439,009A  
 ; FILING DATE: 11-MAY-1995  
 ; CLASSIFICATION: 435  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Casuto, Dianne  
 ; REGISTRATION NUMBER: 40,943  
 ; REFERENCE/DOCKET NUMBER: 4952.US.D1  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 847-938-3137  
 ; INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:  
 LENGTH: 3491 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 US-08-439-009A-2

Query Match 4.1%; Score 105.5; DB 3; Length 3491;  
 Best Local Similarity 21.6%; Pred. No. 0.93;  
 Matches 117; Conservative 60; Mismatches 213; Indels 131; Gaps 25;

10 AAVSRGSSCRARVAPAPAPLAATVVALTLTPA-----RLGNVCAAAAAPAAAPL 65  
 405 ASTGRHRTCCERLLA-----VVGNETAALAGREADAETRELGIDSLVAOLRAK--- 456  
 66 SHVQOAL---AELAKPKDDPTRKHKVCQVAPAVARAIAETG----- 104  
 457 -VSAALIGREVNIALYHPTTPRALAEALAACTEVAQNETARFARNEAAPGPVAVAMAC 514  
 105 -LAPGATTPKQ---LAEGLRRL-----GFDEVFTLFGADLTMEGSELLRLTEH 152  
 515 RLPGGVSTPEEFWELLSGRDAVAGLPTDRGMD--LDSLFRPDP--RSGT----- 561  
 153 LEAHPHSDPELPMFTSCCRGWIAMLEKSYPLLIPVSSCKSPQMLAAMVASYLAERKGI 212  
 562 --AHORGCGFLTEATFADPAFFGM-----SPREALAVDPQOURLMELSMEEVLERAGI 611  
 213 APKDMVAVSIMPCTRKQSEADRDWFCVDADPTLRQLDHYITTVELGNIFKRGINLAELP 272  
 612 PPTSL-----QASP-----IGVYGLIPDGYRPLAE-- 638  
 273 EGEDNPMGVSGAGVLEGTTCGMEALRTAYELFTGTPRLSLSEV----- 321  
 639 -----GGEVEGYLMTGITTSV--ASGRITAYTL--GLEGPALISVDYACSSSLVAVHLA 687  
 322 -----RGMDGIKET-NITMVPAQSKFEELKHAAAARAEAAAGTGGPLAMDGCAFTS 375  
 688 COSLRGESSLAAMGGVTPMPTGMLVDFSRMNSLADGRCKARSA-----GANFGM 740  
 376 EDGGGIGTLRVAVANGLNAAKLTITKMOAGEAKYDFEYI--MACPAGCVGGGGOPRSYDKA 434  
 741 AEGAGMLLE-RLSDARRNGHPVLAIRGTAVNSDGSNGLSNAP-----NGRAQYRVYQQA 795  
 435 ITQR--KQALYINDEKSTLRSHEPNSIRELYDTY--LGEF--LGKHAHELLHTHYVA 487  
 796 LAESGLPADIDAVEAHGTGRLGDPTEARALFEAYGRDREOPHLGYSVSNLGHQAAA 855  
 488 G 488  
 856 G 856

RESULT 6  
 US-09-036-987A-4

Sequence 4, Application US/09036987A  
 Patent No. 6143526

GENERAL INFORMATION:

APPLICANT: Baltz, Richard H.  
 APPLICANT: Broughton, Mary C.  
 APPLICANT: Crawford, Kathryn P.  
 APPLICANT: Madduri, Krishnamurthy  
 APPLICANT: Merlo, Donald J.  
 APPLICANT: Treadway, Paul J.  
 APPLICANT: Turner, Jan R.  
 APPLICANT: Waldron, Clive  
 TITLE OF INVENTION: Biosynthetic Genes For Spinosyn Insecticide  
 TITLE OF INVENTION: Production  
 NUMBER OF SEQUENCES: 39  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Dow Agrosciences LLC Patent Department  
 STREET: 9330 Zionsville Road  
 CITY: Indianapolis  
 STATE: Indiana

COUNTRY: USA  
 ZIP: 46268  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/09/036, 987A  
 FILING DATE: 09-MAR-1998  
 CLASSIFICATION: 435  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Stuart, Donald R.  
 REGISTRATION NUMBER: 28,479  
 REFERENCE/DOCKET NUMBER: 50,608  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (317)337-4816  
 TELEFAX: (317)337-4847  
 INFORMATION FOR SEQ ID NO: 4:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 3170 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 US-09-036-987A-4

Query Match 3.9%; Score 100.5; DB 4; Length 3170;  
 Best Local Similarity 21.3%; Pred. No. 2.6;  
 Matches 116; Conservative 64; Mismatches 190; Indels 175; Gaps 33;

43 EAPARRLGNVACAAAPAAAPAPLSHVQOALAEIAPKDDPTRKH-----VCQVAPAVAR 97  
 1433 EVPAARARAMPANAPAEFGG--SPFARNLAELEPAQ--RRHELVDLYCAOVATV--- 1481  
 98 AIAETGLACATTPQALAEGLRLGFDEVFTLFGADL-----TIME-- 140  
 1482 -----LGHSREEVQPERAFALG---FDSLMAVVDLNRLLTATGRLPTTYFDLP 1530  
 141 EGSELHLRLTEHL-----EAHPHSDPELPMFTSC----- 170  
 1531 NPALAAHLEELVGVASAAVTAASAPASDEPIAIVANSCRRPGAHSPEDLMRLVAG 1590  
 171 -----PGWIAMLEKSY-PDLIPVSSCKSPQMLAAM-----VKSYLEKKGTA 213  
 1591 TEVIGFEPDGRMDA--EGLYDP-----ASRPTTYARMAAGLYDAGEFADLFGIS 1641  
 214 PKDMVAVSIMPCTRKQSEADRDWFCVD--ADPTLRQLDHYITTVELGNIFKRG--INL 268  
 1642 PRELAMD-----POQRLVLEIAWEALERAGIDPLSLKSGGVGYIGAGS--RGYATDV 1693  
 269 AELPEGEDNPMGVSGAGVLEGTTCGMEALRTAYEL-FTGTPLP-----R 315  
 1694 RQPF--EAEGYILTGTSASVLSG-----RVAVSFGFEPGPAVTVDTCSSSLVALH 1742  
 316 LLSLEYRG--MDGIKETNITMVPAQSKFEELKHRAA---RAEAAAGTGGPLAMDGG 370  
 1743 LACOSLRSGCDLALAGVTVSTP--EMVEVSROGLAPDGRCKAFASAG--TGNGSG 1800  
 371 AGTSDGGRGIGTLRVAVANGLNAAKLTITKMOAGEAKYDFEYI--MACPAGCVGGGGOPR 429  
 1801 A-----GLLERLSDARRNGHRVLAIVRGSAVNODGASNGLAIP-----NGPSOOR 1847  
 430 STDKAITQRQAALVNLDEKSTLRSHEPNSIRELYDTYLGEPGLGHKAHELLHTHYVAG 489  
 1848 VTNQALAA--NAALSVDVAV--BAHGT-----GTRLGDPTE--EAQALATYV--GQ 1890  
 490 VEKED 494  
 1891 ARERD 1895

RESULT 7  
 US-09-0370-700-4

```

; Sequence 4, Application US/09370700
; Patent No. 6274350
; GENERAL INFORMATION:
; APPLICANT: Balcz, Richard H
; APPLICANT: Broughton, Mary C
; APPLICANT: Crawford, Kathryn P
; APPLICANT: Madduri, Krishnamurthy
; APPLICANT: Treadway, Patti J
; APPLICANT: Turner, Jan R
; APPLICANT: Waldron, Clive
; TITLE OF INVENTION: Biosynthetic Genes For Spinosyn Insecticide
; FILE REFERENCE: 50489 Div1
; CURRENT APPLICATION NUMBER: US/09/370,700
; EARLIER FILING DATE: 1999-08-09
; EARLIER APPLICATION NUMBER: US 09/36987
; NUMBER OF SEQ ID NOS: 39
; SOFTWARE: Patentl Ver. 2.0
; SEQ ID NO 4
; LENGTH: 3170
; TYPE: PRT
; ORGANISM: Saccharopolyspora spinosa
US-09-370-700-4

```

```

Query Match      3.9%; Score 100.5; DB 4; Length 3170;
Best Local Similarity 21.3%; Pred. No. 2.6;
Matches 116; Conservative 64; Mismatches 190; Indels 175; Gaps 33;

```

```

OY 43 EAPARRLGNVACAAAPAAEAPLSHVQALAEAKPKDDPRKH-----VCQVAPAVRV 97
DB 1433 EVPAARKAMPANGPAPGCG-----SPFARMLAELEPAQ-----RRHETLVLCVQAVTV--- 1481
OY 98 AIAETLGLAPATTKQALEGRRLGFDEVEDTLFGAL-----TIME-- 140
DB 1482 -----LGHGREVQPERAFRAIG-----FDSLAAVDLRNLTATGLRPTTVETVP 1530
OY 141 EGSELLHRLTEHL-----EAHPHSDPELPMFTSCC----- 170
DB 1531 NPAALAAHLEELVDVSAAVTASAPASDEPRIVAMSCFPGASHPEDMLRLVAAG 1590
OY 171 -----PGWIMLEKST-PDLIPYVSSCKSPOMLAAM-----VKSTLAEEKGIA 213
DB 1591 TEVIGEPSPDRGMWA--EGLYDPD-----ASRPCTYARWAGFLYDAGEFDADLFGIS 1641
OY 214 PKDWMVSIIMPCTRKQSEADRMFCVD--ADPTLROLDHVTYTELGNIFEREG--INL 268
DB 1642 PREALAMD-----PQORLVLEIAEALERAAGIDPLSKSGVGTIIGAGS---RGYATDV 1693
OY 269 AELPGEWDMPMGVSGAGVLEGTGGVMEALRTAYEL-FYGTPLP-----R 315
DB 1694 ROFPE-EAEGYILIGTSAVSG-----RAVSRGFEGPRAVTVDTACSSSLVALH 1742
OY 316 LSLSEVRG--MDGIKETNITWVPAPGSKFEELKRRAAA---RAEAAHGTTPGLAMDGG 370
DB 1743 LACQSLRSGECDLALAGVTVWSTP-EMFVEFSRORGLAPDGRCKSFESADG-TGMEEG 1800
OY 371 AGFTSEDRGGITLVAANGNNAKLLITKQACEAKYDVEI-MACPRAGCVGGCGQPR 429
DB 1801 A-----GLLLERLSDAHNRHGRVLAAYVRSAGAVNODGASNGLAAP---NGPSQOR 1847
OY 430 STDKATIKORQALATNLDEKSTLRSHENPSIRELYDYTLGEPHKKAHELLHTHYVAGG 489
DB 1848 VTNQALA---NAALSASVDVAV--EAHGT-----GTRLSGPI--EAOALATATY---CQ 1890
OY 490 VEKED 494
DB 1891 ARERD 1895

```

```

RESULT 8
US-08-426-125-4
; Sequence 4, Application US/08426125
; Patent No. 5853973

```

```

; GENERAL INFORMATION:
; APPLICANT: Kakefuda, Genichi
; APPLICANT: Ott, Karl-Heinz
; APPLICANT: Kwagh, Jae-Gyu
; APPLICANT: Stockton, Gerald W.
; TITLE OF INVENTION: Structure-Based Designed Herbicide
; TITLE OF INVENTION: Resistant Products
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSER: Darby & Darby
; STREET: 805 Third Avenue
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10022-7513
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentl Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/426,125
; FILING DATE: 20-APR-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Robinson, Joseph
; REGISTRATION NUMBER: 33,448
; REFERENCE/DOCKET NUMBER: 0646/0A674
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212)-527-7783
; TELEFAX: (212)-753-6237
; TELEX: 236687
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 638 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEICAL: NO
; ORIGINAL SOURCE:
; ORGANISM: Zea mays
US-08-426-125-4

```

```

Query Match      3.9%; Score 99.5; DB 2; Length 638;
Best Local Similarity 22.3%; Pred. No. 0.23;
Matches 113; Conservative 56; Mismatches 219; Indels 119; Gaps 23;

```

```

OY 39 LATEAPARRLGNVACAAAPAAEAPLSHVQALAEAKPKDDPRKHVCQVAPAVRVA 98
DB 1 MATAAASATLTGATTAPKARRRHILATRALAA-----PIR---CSAASFAMPRA 50
OY 99 IAEITGLAP-GATTPKO---LAEGRLRGFDEVEDTLFGADLTMEGSELLHR--LT 150
DB 51 PRAT-PLRPMGPTDRKAGDILVESLERGVADVAPYFGASMEI---HQALTRSPVIA 105
OY 151 EHLEAHPSDELPHTSCPCGWIMLEKSYDILIPYVSSCKSPOMLAAMKSTLAER- 209
DB 106 NLEFRHEGE-----AFASGVYRSGRGVGCATSGPGATNIVSALADL 151
OY 210 -----KGIAPRDV-----WVSIMPCTRKQSEADRMFCVDADPTLRQLDHVIT 254
DB 152 LDSVWMAITGVPRRMICTDAFOETPIVEYR--STIKHNLVLDVDDIPRVQGEAPFL 209
OY 255 VELGNIFERGINLAELPEG-----EWDNPMGVSGAGVLEGTGGVMEALRTAYE 306
DB 210 ASSG-----RPGPVLDIPKDIQOQMAVPMWDMKPMGL-----PGYIARLPRPATE 255
OY 307 LFTGTPRLSLSEVRGMDGIKETNITW---PARGSKFEELKRRAAAREAAHGT 362
DB 256 L-----LEQVLRIVGESRRPVLYVGGCAASG---EEL-----RPFVELTGI 295
OY 363 GLAMDGAGGFTSEDG-----RGGITLRVAV-----ANGIGNAKKLLITKQAGEA 407

```

Db 296 VTTTLGLGNFSDDDLRLMLGMHCTVYANYAVDKADLLALGVFDRVGTKEIAFAS 355  
QY 408 KYDFEIMACPGACVCGCGPRSTDKAITOKRQALYNDEKSTLRSHENPSIRELYDT 467  
Db 356 RAKIVHVDIDPAE-IGKNQPHYSICADVKALQGNALLBSTSKSPFSGWDELQO 414  
QY 468 YLGE-PLGHR-AHELLHTHYVAGVEE 492  
Db 415 OKREFPLGYKTSNEEIOPOYAIQVLD 441

RESULT 9  
US-08-455-355-4  
Sequence 4, Application US/08455355  
Patent No. 5928937  
GENERAL INFORMATION:  
APPLICANT: Kakefuda, Genichi  
APPLICANT: Otc, Karl-Heinz  
APPLICANT: Kuegh, Jae-Gyu  
APPLICANT: Stockton, Gerald W.  
TITLE OF INVENTION: Structure-Based Designed Herbicide  
TITLE OF INVENTION: Resistant Products  
NUMBER OF SEQUENCES: 10  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: Darryl A. Darby  
STREET: 805 Third Avenue  
CITY: New York  
STATE: New York  
COUNTRY: USA  
ZIP: 10022-7513  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/455,355  
FILING DATE: 31-MAY-1995  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Robinson, Joseph  
REGISTRATION NUMBER: 33,448  
REFERENCE/DOCKET NUMBER: 0646/1A674-US1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212)-527-7783  
TELEFAX: (212)-753-6237  
TELEX: 236687  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 638 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: NO  
ORIGINAL SOURCE:  
ORGANISM: zea mays  
US-08-455-355-4

Query Match 3.9%; Score 99.5; DB 2; Length 638;  
Best Local Similarity 22.3%; Pred. No. 0.23;  
Matches 113; Conservative 56; Mismatches 219; Indels 119; Gaps 23;

QY 39 LATLEPARLGNVACAAAPAAEAPLSHVQALAEAKKDPTRKHVCVQVAPAVRA.98  
Db 1 MATAAASTALTGATTAAPARRARRHLATRALAA-----PIR---CSAASPMPMA 50  
QY 99 IAEITGLAP-GATTPKO---LAEGLRRLGDFEVDFTLFGADLTINEGSELLHR---LT 150  
Db 51 PPT-PLRTPGFPDPRKAGDILVESLERCYRDVFAIPGASMET-----HQLTRSPVIA 105  
QY 151 EHLEAPHSDEPLPMFTSCPGWIAMLEKSYPLIPVSSCKSPOMLAAMVKSYLEAK- 209

Db 106 NHLFRHDECE-----AFASGYARRSGRGVCATISCPGATNLVSAADAL 151  
QY 210 -----KGIAPKDMY-----MVISMPCTRKQSEADRMPCVADADTLRLQDHLVIT 254  
Db 152 LDSVPVAVITGOVPRMIGTDAPOETPIVEYR--STHNTLVLDVDDIPRYVQDAFPL 209  
QY 255 VELGNIFKRGINLAELPEG-----EMDPMGVSGAGVLTGTGCVMEALATAYE 306  
Db 210 ASSG-----RPGVVLVDIPDIOQMAVPYWDKPMSL-----PGVIALPKPPAPE 255  
QY 307 LFTGTPPLRLSLSEVKGMDIKETNTMY----PAPGSFEEELKRRAAAREAAAHGP 362  
Db 256 L-----LEQVRLVGESRRPVLVYGCGCAASG---EBL-----RRFVELTGIP 295  
QY 363 GPLANDGAGFTSEDC-----RGITLRLAV-----ANGIGNAKKLLTKMOAGRA 407  
Db 296 VTTTLGLGNFSDDDLRLMLGMHCTVYANYAVDKADLLALGVFDRVGTKEIAFAS 355  
QY 408 KYDFEIMACPGACVCGCGPRSTDKAITOKRQALYNDEKSTLRSHENPSIRELYDT 467  
Db 356 RAKIVHVDIDPAE-IGKNQPHYSICADVKALQGNALLBSTSKSPFSGWDELQO 414  
QY 468 YLGE-PLGHR-AHELLHTHYVAGVEE 492  
Db 415 OKREFPLGYKTSNEEIOPOYAIQVLD 441

RESULT 10  
US-08-804-227C-10  
Sequence 10, Application US/08804227C  
Patent No. 5876991  
GENERAL INFORMATION:  
APPLICANT: DeHoff, Bradley S.  
APPLICANT: Kuesters, Stuart A.  
APPLICANT: Kosteck, Paul R., Jr.  
APPLICANT: Sutton, Kimberly L.  
TITLE OF INVENTION: POLYKETIDE SYNTHASE GENES  
NUMBER OF SEQUENCES: 15  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: THOMAS G. PLANT 1501  
STREET: LILLY CORPORATE CENTER  
CITY: INDIANAPOLIS  
STATE: IN  
COUNTRY: USA  
ZIP: 46285  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: MS-DOS  
SOFTWARE: ASCII(DOS) Text only  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/804,227C  
FILING DATE: February 21, 1997  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Plant, Thomas, G.  
REGISTRATION NUMBER: 35,784  
REFERENCE/DOCKET NUMBER: X-8231  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 317-276-2459  
INFORMATION FOR SEQ ID NO: 10:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3724 amino acids  
TYPE: amino acid  
TOPOLOGY: unknown  
MOLECULE TYPE: peptide  
US-08-804-227C-10

Query Match 3.9%; Score 99.5; DB 2; Length 3724;  
Best Local Similarity 22.8%; Pred. No. 4.4; 255; Indels 111; Gaps 22;  
Matches 124; Conservative 54; Mismatches 111; Gaps 22;

```

QY 15 RGSSCRARQVAPRPLAASITRVVALATEAPARRLGNVACAAAAPA-AEAPLSHQOLA 73
Db 1382 RGSACGAVADVENSREFAASTWTPVPAVLEFDIPYQIRRAELAPSPGDSITSELVRLT 1441
QY 74 ELAKPKDDPTREKHVCQVAVPAVRVAIETGLIAGATTPKOLAGELRLRGDE----- 126
Db 1442 AQSGHKRHAT-----LRLRVRAHAAVAVLGSSGDVSS--AAAFRLGSDLSLAELR 1492
QY 127 -----VEDTLEFGADULTIMEGSELLHRLTEHLEAHP-----HSDEPLMET 167
Db 1493 DRLESTGTGLKPTSLVLEPHSSPALA--RHLEGEELLGRNDTADRAGPDPPTVPTDEPIAIG 1551
QY 168 SCC--PQMIAMLESYDPLIPYVSSCKSPQKMLAAMKSYLAEEKGIAPKDMVWSIMPC 225
Db 1552 MACRLPGVOSPEDLW-DLITGGTDALTPFPTRNGMDNETLYDDPDPSPGHHTYVREGCG 1610
QY 226 TRKQSEADRDQFCVD-----ADPTLRLDHYITVTEIGNIFKERGIJLAELPEGEDNP 279
Db 1611 LHDAEEDPQFGFGISPREALAMDPOQRIL-LETSWE--SFERAGIDPVELR----- 1658
QY 280 MGVSAGAGVLEFGTTGGVMEALRFAVELEF-----TGTPPLRLS--LSEVRGMDGKETNI 332
Db 1659 ---GSRGVGFVGTNGOHVPLLOQDDENFEDGYIATGNSASVMSGRLSYFEGLEG---PAV 1712
QY 333 TMVPAPSKEPEEL-LKHRAARAE-----AAAHGTPGLANDGAGTFSEOGRGIT 383
Db 1713 TVDTRPSSLSIALHLIAVOSLRGRCDDYALAGATVMSPEMLVEPAROAVSPDRSKAF 1772
QY 384 LRVAVANGIGNAKRLITFQMGAEKRYDVEIIMACPAGV-----GGCGGPRS 430
Db 1773 AEADQVGLAEGAGMLVERISEAKQKHPIYAVVRGSAVNOGASNGLTAPSGPAQQRV 1833
QY 431 TDKAITQKQALYNLDEKSTLRSSHENPSIRELYDTYLGSPLGKRAHELLHTHYVAGV 490
Db 1833 IREALA--DAGLTPADVADVAEAGHTGP-----LGDPIL-EGALLATY---GR 1874
QY 491 EEKD 494
Db 1875 DRRD 1878

RESULT 11
US-08-804-198-4
Sequence 4, Application US/08804198
Patent No. 5945320
GENERAL INFORMATION:
APPLICANT: Burgett, Stanley G.
APPLICANT: Kuhstoss, Stuart A.
APPLICANT: Rao, Nagaratja R.
APPLICANT: Richardson, Mark A.
APPLICANT: Rostek, Paul R., Jr.
TITLE OF INVENTION: PLATENOLIDE SYNTHASE GENE
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: PAUL R. CANTRELL 1138
STREET: LILLY CORPORATE CENTER
CITY: INDIANAPOLIS
STATE: IN
COUNTRY: USA
ZIP: 46285
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: Macintosh
OPERATING SYSTEM: Macintosh 7.0
SOFTWARE: Microsoft Word 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/804,198
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: CANTRELL, PAUL R.
REGISTRATION NUMBER: 36,470
REFERENCE/DOCKET NUMBER: P9113

```

```

: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 317-276-3885
: INFORMATION FOR SEQ ID NO: 4:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 3724 amino acids
: TYPE: amino acid
: TOPOLOGY: unknown
: MOLECULE TYPE: peptide
: US-08-804-198-4

Query Match      3.9%; Score 99.5; DB 2; Length 3724;
Best Local Similarity 22.8%; Pred. No. 4.4;
Matches 124; Conservative 54; Mismatches 255; Indels 111; Gaps 22;

   15  RGSSCRARQVAPRAPLAATVRAVALTLEAPARRIGVACAAAPA-AEAPLSHYQALA 73
      ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db    1382  RGSACGAVADVEMSRFASATYTWBPAYVFDIDIPVQRLEAAELPSTGDSITSELVREL 1441
      74  ELARPKDDPTKKHYCVQVAPVRAVALETGLACGATTPKRLAGLRRLGDE----- 126
      1442  AOSGKHRRAT-----LLRLVRAHAAAVLIGSSGDAAVS--ARAFLRGDSITALELR 14922
      127  -----VFDTLFGADLTIMEEGSELLRLTHELEAHP-----HSDEPLPMFT 167
      1493  DRLSTGCLKLPTLSYLFPHSSPALA-RHLGDELLGRNDTADRCAGPTFPVTFDEIALIG 15515
      168  SCC--PGMIAMLEKSYPDLPYVSSCKSPOMMLAAMKSYLAEEKGIAPKDMVNVSI MPC 225
      1552  MACRLPGVGQSPEDLM-DLTGTDATLPFPTRNGMDNETLYDDPDPSGCHHTYVREGGF 16100
      226  TRKQSEADRMFCVD-----ADPLRQJDHYITVELGNTFKKNGIIMALELPEGEHNP 279
      1611  LHDAAEFPDGFEGGISPREALAMPQQRLL-LETSWE--SFERAGIDPVELR----- 1658
      280  MGVGSGAGVLTCTGGCVMEALRTAYELF-----TGTPLEPLRS--LSEVRGMDGKETNI 332
      1659  ---GSRTGVFPGTNGQNHVPLRLDQDDENFDEGTIATGASVMSGSLSTVFGLEG---P 17112
      333  TMVPAPGSKFEEL-LKHRAAARAE-----AAAHGTGPRLAMDGAGFTSEDRGGIT 383
      1713  TVDTACASIALAHLVALVQSLRRGECDYALAGGATFVSPMLVEFARORAVSPDGRSAF 1772
      384  LRVAVANGIANKKLITTMQAGEAKYDVEVELMACAGCV-----GGCGQPRS 430
      1773  AEADGVVLEAGAGMLIVERLSEADKKGHPLYAVVRSVAVNODGASNGLTAPSPGAQQRV 1832
      431  TDKATTKROALYNLDEKSTLRSHENPSTRELDTYLGEPILGKAHELLHTHYVAGV 490
      1833  IREALA---DAGLPADPADVAEAGTGTP-----LGDPY--EAGALATVY---GR 1874
      491  EEKD 494
      : |
      1875  DRRD 1878

RESULT 12
US-08-804-227C-8
: Sequence 8, Application US/08804227C
: Patent No. 5876991
: GENERAL INFORMATION:
: APPLICANT: DeHoff, Bradley S.
: APPLICANT: Kuhstoss, Stuart A.
: APPLICANT: Rostock, Paul R., Jr.
: APPLICANT: Sutton, Kimberly L.
: TITLE OF INVENTION: POLYKETIDE SYNTHASE GENES
: NUMBER OF SEQUENCES: 15
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: THOMAS G. PLANT 1501
: STREET: LILLY CORPORATE CENTER
: CITY: INDIANAPOLIS
: STATE: IN
: COUNTRY: USA
: ZIP: 46285

```

COMPUTER READABLE FORM:  
MEDIUM TYPE: FLOPPY disk  
OPERATING SYSTEM: MS-DOS  
SOFTWARE: ASCII(DOS) text only  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/804,227C  
FILING DATE: February 21, 1997  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Plant, Thomas, G.  
REGISTRATION NUMBER: 35,784  
REFERENCE/DOCKET NUMBER: X-8231  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 317-276-2459  
INFORMATION FOR SEQ ID NO: 8:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 4550 amino acids  
TYPE: amino acid  
TOPOLOGY: unknown  
MOLECULE TYPE: peptide  
US-08-804-227C-8

Query Match 3.9%; Score 99.5; DB 2; Length 4550;  
Best Local Similarity 18.5%; Pred. No. 6.2;  
Matches 114; Conservative 62; Mismatches 192; Indels 247; Gaps 27;

QY 2 SALVLRPCAAVSRGSS--CRARQA-----PRAPLAATVVALATLEAPARR---- 48  
DB 233 AAVVLRPLDALDADDPYCVVGVAVGNDGGP-----GLTADREGQEA 278  
QY 49 LGNVACAA-----APAEAPLSHVQALAEIAEKRPDP----- 82  
DB 279 VLRAACQAQAVDPAEYFVELHGTGTPVGDPEAHALGAHSGGRAPDDPLVGSVKTNI 338  
QY 83 -----TRKHCY-----QVAPAVVAIAETLGLAPGA 109  
DB 339 GHLEGAAGIAGLVKALCLERETLPGLSINATPSPAIPDLQILKQVOTAAAE-LPLAPGG 397  
QY 110 TTPKOALAEIRLRGFDEVEFTLFGADLTMEGSELHRTLEHLEAHPHSDPELPMFTSC 169  
DB 398 A---PLLAGVSSSGI-----GGINCHVYLEHLPSPRP----- 427  
QY 170 CPWMIALEKSYDPLIPYSSCKSPQMLLAAYK-STLAEEKGIAPKDYVW----- 220  
DB 428 ----AVSAASLPDVPPLLSARSEGALRAQAVRLGEYVERVGADPRDVAVSLASTRTLE 483  
QY 221 ---SIMPCTRKQS-----EADR-----DW----- 236  
DB 484 EHRAVVPCCGGRGELVVALGGFAAGRVSGVRSRAVPGVGLFTGGAGVGMKGGLYA 543  
QY 237 ----FCVDADPTLRDLHV-----ITVVELGNIKEKGINLAELPEGEMDPMKVGSGAGV 288  
DB 544 GGGVFAEVLDEVLSMGEVDGRSLRDVMDVDVADAGAG-----DAGAGAGAGVGS 595  
QY 289 LEFTTGGV-----MEALRTAYE-----LFTGTPPLPLSLSEYRGMDGK 328  
DB 596 GSGSVGLGRTPEAQPALFALEVALFRALEARGVEVSVLGHSGVEVAAYAVGLSG 655  
QY 329 ETITMTPAPGSKF-----PELLKRAAARAE-----AAAGTGGPLAMDG 369  
DB 656 DA-VRLVVARGLMGLIPVGGGMKSVGASVYRGVEGLGEWVSAVAANG-PRSVYLSG 713  
QY 370 GAG-----FTSEDRGRTLRVAVANG-----LGNKKLITKMQAGEARYDFVEIM 415  
DB 714 DVGLEBSVASLMDGDVECRRLDVSIGFHSVLMPEVLGERRGVYESLEFRVAPGVVVS 773  
QY 416 ACPAGCVGG--GQP 428  
DB 774 GVGSGVVGSGELGDP 788

RESULT 13

US-08-804-198-2  
Sequence 2, Application US/08804198  
Patent No. 5945320  
GENERAL INFORMATION:  
APPLICANT: Burgett, Stanley G.  
APPLICANT: Kuhstoss, Stuart A.  
APPLICANT: Rao, Nagaraja R.  
APPLICANT: Richardson, Mark A.  
APPLICANT: Roostek, Paul R., Jr.  
TITLE OF INVENTION: PLATENOIDE SYNTHASE GENE  
NUMBER OF SEQUENCES: 6  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: PAUL R. CANTRELL, 1138  
STREET: LILLY CORPORATE CENTER  
CITY: INDIANAPOLIS  
STATE: IN  
COUNTRY: USA  
ZIP: 46285  
COMPUTER READABLE FORM:  
MEDIUM TYPE: FLOPPY disk  
OPERATING SYSTEM: Macintosh  
SOFTWARE: Microsoft Word 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/804,198  
FILING DATE:  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: CANTRELL, PAUL R.  
REGISTRATION NUMBER: 36,470  
REFERENCE/DOCKET NUMBER: P9113  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 317-276-3885  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 4550 amino acids  
TYPE: amino acid  
TOPOLOGY: unknown  
MOLECULE TYPE: peptide  
US-08-804-198-2

Query Match 3.9%; Score 99.5; DB 2; Length 4550;  
Best Local Similarity 18.5%; Pred. No. 6.2;  
Matches 114; Conservative 62; Mismatches 192; Indels 247; Gaps 27;

QY 2 SALVLRPCAAVSRGSS--CRARQA-----PRAPLAATVVALATLEAPARR---- 48  
DB 233 AAVVLRPLDALDADDPYCVVGVAVGNDGGP-----GLTADREGQEA 278  
QY 49 LGNVACAA-----APAEAPLSHVQALAEIAEKRPDP----- 82  
DB 279 VLRAACQAQAVDPAEYFVELHGTGTPVGDPEAHALGAHSGGRAPDDPLVGSVKTNI 338  
QY 83 -----TRKHCY-----QVAPAVVAIAETLGLAPGA 109  
DB 339 GHLEGAAGIAGLVKALCLERETLPGLSINATPSPAIPDLQILKQVOTAAAE-LPLAPGG 397  
QY 110 TTPKOALAEIRLRGFDEVEFTLFGADLTMEGSELHRTLEHLEAHPHSDPELPMFTSC 169  
DB 398 A---PLLAGVSSSGI-----GGINCHVYLEHLPSPRP----- 427  
QY 170 CPWMIALEKSYDPLIPYSSCKSPQMLLAAYK-STLAEEKGIAPKDYVW----- 220  
DB 428 ----AVSAASLPDVPPLLSARSEGALRAQAVRLGEYVERVGADPRDVAVSLASTRTLE 483  
QY 221 ---SIMPCTRKQS-----EADR-----DW----- 236  
DB 484 EHRAVVPCCGGRGELVVALGGFAAGRVSGVRSRAVPGVGLFTGGAGVGMKGGLYA 543  
QY 237 ----FCVDADPTLRDLHV-----ITVVELGNIKEKGINLAELPEGEMDPMKVGSGAGV 288  
DB 544 GGGVFAEVLDEVLSMGEVDGRSLRDVMDVDVADAGAG-----DAGAGAGAGVGS 595



Search completed: June 3, 2003, 16:09:43  
Job time : 32 secs

---



Db 4600 EEPFLALAGL-RVTDLPAGETPAK-----VDLDLTLHEVAGRDGKHATL----- 4641  
QY 266 INLAELPEGEHNDPMKGVSGAGVLEGTGGVMEALRTAYELFTGTPRLRLSLSEVRGD 325  
Db 4642 LGAADLEFDEB-----TVRLADRLRT-LEMAAAPDRDLDEVLSPG 4684  
QY 326 G-----JKEFTITVPAPGSKFEELKHHRAAARAAAHGTPGLWDGAGFTSEGRG 381  
Db 4685 ERSRLLEWMDTRAPVESVPRLLFAEQVAAPDAVAVGEG-VSMT-----YRELDARSD 4739  
QY 382 ITLRVAVANGLGNAKLLITMKG-----AG-----EAKYDVEINA 416  
Db 4740 ALARSLVAAAGVGVESPPVVALERSPEVLSAFLAVAKAGVFPVDSLMPARVDAV-VAD 4798  
QY 417 CPAG-----CYGGGGOPRSTOKAITOKRQ-----ALVNLDEKSTL 452  
Db 4799 CGARLAVADRPMSGLTVASAGLSDSAVVSGLDADRAVVLPAQVPAAVYRMVYSGST 4858  
QY 453 RR-----SHEN-----PSIRELYD----- 466  
Db 4859 GRKGVVTHQNLVDLATDTCMGPTPRVFLPHAPHADASSYELIWPILLNGTVVAPRRS 4918  
QY 467 ---TYLGEPLGHAHELLHTHYVAG 488  
Db 4919 IDATVLRDLIG--AHELTHVHTAG 4941

RESULT 2  
US-09-737-149-4  
; Sequence 4, Application US/09737149  
; Patent No. US20020077466A1  
; GENERAL INFORMATION:  
; APPLICANT: Spaderna, Steven K  
; APPLICANT: Quinn, Kerry E.  
; APPLICANT: Shimkets, Richard A.  
; APPLICANT: Muralidhara, Padigaru  
; APPLICANT: Spytek, Kimberly A.  
; TITLE OF INVENTION: Polypeptides and Nucleic Acids Encoding Same  
; FILE REFERENCE: 15966-620 CIP  
; CURRENT APPLICATION NUMBER: US/09/737,149  
; CURRENT FILING DATE: 2001-06-15  
; PRIOR APPLICATION NUMBER: 60/170,564  
; PRIOR FILING DATE: 1999-12-14  
; PRIOR APPLICATION NUMBER: 60/173,165  
; PRIOR FILING DATE: 1999-12-27  
; PRIOR APPLICATION NUMBER: 60/173,362  
; PRIOR FILING DATE: 1999-12-27  
; PRIOR APPLICATION NUMBER: 60/174,404  
; PRIOR FILING DATE: 1999-12-29  
; PRIOR APPLICATION NUMBER: 60/174,962  
; PRIOR FILING DATE: 2000-01-07  
; PRIOR APPLICATION NUMBER: 60/223,929  
; PRIOR FILING DATE: 2000-08-09  
; NUMBER OF SEQ ID NOS: 49  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 4  
; LENGTH: 965  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-737-149-4

Query Match 4.7%; Score 122; DB 10; Length 965;  
Best Local Similarity 21.9%; Pred. No. 0.1;  
Matches 120; Conservative 46; Mismatches 189; Indels 192; Gaps 22;

QY 7 KRCVAVSRGSSCARQVAPAPLAASVRYALATLTPAPRRIGNVA--CAAAPAAEAP 64  
Db 25 QPCGVLLARLGGSVRLGALLPAPLARAARALALAPLPINLSLELVAAAPPRDP 84  
QY 65 LSHVQ---QAL-----AELAKPKDDP--TRKHVCQVAVAPVAVVAIETGLAP-GATTP 112

Db 85 ASITRGLCOALVPPGVAALLAFPEARPELLQHLFAAATETPVLSLRREARAPLGAAPN 144  
QY 113 KQALBGLRLGDEVEDTLFGADLTIMBGSSELHRLTLEHLEAHPHSDPRLPMFTSCCPG 172  
Db 145 FHL-----OLHWASPLET-----LLDVLVAVLQHNHAMEDGLACRTQDGC 185  
QY 173 WIMLEKSY-----PDLIPVSSCKSPOMLAAMVKSYLEKKGIAFKDMVNVIMPCTRK 228  
Db 186 GLVALMTSSAGRRPQLVULDSRDTGDAGLRARLAPMAAPVSGEAP----- 231  
QY 229 QSEADNDWFCVADPRLRLQDHYITTYELG-NIFKRGGINLAELPGEHNDPMKGVSGAG 287  
Db 232 -----VPAVLLGCDIARRRVLEAVPPGPHW----- 258  
QY 288 VLEFTGGVMEALRTAYELFTGTPRLRLSLSEVR-----GMDGIR 328  
Db 259 -LIGTP--LPPKALPFA-----GLPGLALSGVAPRPLEAHTDVQVVARALSSAAV 310  
QY 329 ETNITVPAP-----GSKFEELKHHRAAARAAAHGTPGLWDGAGFTSEGRG 381  
Db 311 QPKRALLPAPVNGDLOPAGPESPGRFLAFLANTSFQGRTPV-WYTGSS-PDEDGQ-- 366  
QY 382 ITLRVAVANGLGNAKLLITMKGAEAKYDVEIMACPAQ--CYGGGGOPRSTOKAITOKR 439  
Db 367 -----CPAGQLCL-----DPGTNDSATVLDAL 387  
QY 440 QALVNLDEKSTLRR-----SHENPSIRELY-----DTYLGE 471  
Db 388 FAALANGAPRALRKCCYGCIDLERLADTFFDEFLYVGKXKYGALBDGHWGTGLVGD 447  
QY 472 PLGKHAH 478  
Db 448 LLAGRAH 454

RESULT 3  
US-09-737-149-6  
; Sequence 6, Application US/09737149  
; Patent No. US20020077466A1  
; GENERAL INFORMATION:  
; APPLICANT: Spaderna, Steven K  
; APPLICANT: Quinn, Kerry E.  
; APPLICANT: Shimkets, Richard A.  
; APPLICANT: Muralidhara, Padigaru  
; APPLICANT: Spytek, Kimberly A.  
; TITLE OF INVENTION: Polypeptides and Nucleic Acids Encoding Same  
; FILE REFERENCE: 15966-620 CIP  
; CURRENT APPLICATION NUMBER: US/09/737,149  
; CURRENT FILING DATE: 2001-06-15  
; PRIOR APPLICATION NUMBER: 60/170,564  
; PRIOR FILING DATE: 1999-12-14  
; PRIOR APPLICATION NUMBER: 60/173,165  
; PRIOR FILING DATE: 1999-12-27  
; PRIOR APPLICATION NUMBER: 60/173,362  
; PRIOR FILING DATE: 1999-12-27  
; PRIOR APPLICATION NUMBER: 60/174,404  
; PRIOR FILING DATE: 2000-01-07  
; PRIOR APPLICATION NUMBER: 60/223,929  
; PRIOR FILING DATE: 2000-08-09  
; NUMBER OF SEQ ID NOS: 49  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 6  
; LENGTH: 971  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-737-149-6

Query Match 4.7%; Score 122; DB 10; Length 971;  
Best Local Similarity 21.9%; Pred. No. 0.1;

	Matches	120: Conservative	46: Mismatches	189: Indels	192: Gaps	22:
OY	7	KPCAASVIRGSSCCAROVAPRAPLAASVYVALATLEAPARRLGNA--CAAAAPAAEAP	64			
Db	25	QPCGVILRIGSGSVRLGALPLPRAPRAARAARALAAALAPRLPHLSLELVAAAPAPDP	84			
OY	65	LSHVQ-----QAL-----AELAKPRDDP--TAKHYCVQVAPAVRVAIAETLGLAP-GATTP	112			
Db	85	ASLTRGLCQALVYPPGVAAALLAFPEARPELLOHFLAAATETPVLSLRRERAPPLGAINP	144			
OY	113	KQLABGRIRLGFDEVFTDLFGADLTIMEGSELLHRLTEHLEAPHPSEDEPLPMTSCCPG	172			
Db	145	FHL-----QLHWASPLET-----LLDVLVAVILQAHMEDVGLACRTQDPG	185			
OY	173	WIAHLEKSY---PDLLPYVSSCKSPOMLAAWKSYLAEEKGLAPKDMVAVSIMPTCTRK	228			
Db	186	GLVALMTSRAGRPOLVLDLSRRDPTDAGLRARLAPMAAPVGGEAP-----	231			
OY	229	QSEADRWFCVDADPTLRQLDHYITTVELG-NIFERGINLAEPEGEMDNMGVSGAG	287			
Db	232	-----VPAVILGCDIARARVLEAVPPGPH-----	258			
OY	288	VLFTGTGVMBAALRTAYELTGTPLPLSLSEVR-----GMDGIRK	328			
Db	259	-LLGTP-LPRKALPTA-----GLPRGLALGEVARPPLLEAIIHDYQLVARALGSAANQ	310			
OY	329	ETNITWMPAP-----GSKFEELIKHPAARAEAAAHGTGPLAMDCAGFTSDEGRKG	381			
Db	311	QPKRALLPAPNCCDLPDPAGESPGRFLARFLANTSFOGRGTPV-WVTGSS-PDEGDO--	366			
OY	382	ITLRVAVANGGNKKLITKQAGEAKYDFEIMACPG--CVGCGGQPRSDKAITQKR	439			
Db	367	-----CPAGQLCL---DPTINDSATLDAL	387			
OY	440	QALYINDEKSTLRR-----SHENPSIRELY-----DTYIGE	471			
Db	388	FAALANGSAPRALRKCQCYCIDILLERLAEDTPDFEELIYVGDGKYKALRDRGTGLVGD	447			
OY	472	PLGKHAH 478				
Db	448	LLAGRAH 454				
RESULT 4						
US-10-106-534-2						
Sequence 2, Application US/10106534						
Patent No. US2002016868A1						
GENERAL INFORMATION:						
APPLICANT: Curtis, Rory A.J.						
TITLE OF INVENTION: 14691, A Human Glutamate Receptor Family						
FILE REFERENCE: MP101-042P1RM						
CURRENT APPLICATION NUMBER: US/10/106.534						
CURRENT FILING DATE: 2002-03-26						
PRIOR APPLICATION NUMBER: 60/279,086						
PRIOR FILING DATE: 2001-03-27						
NUMBER OF SEQ ID NOS: 7						
SOFTWARE: FastSeq for Windows Version 4.0						
SEQ ID NO 2						
LENGTH: 998						
TYPE: PRT						
ORGANISM: homo sapiens						
US-10-106-534-2						
Query Match 4.7%; Score 120; DB 9; Length 998;						
Best Local Similarity 22.3%; Pred. No. 0.16;						
Matches 129; Conservative 45; Mismatches 215; Indels 190; Gaps 24;						
OY	7	KPCAASVIRGSSCCAROVAPRAPLAASVYVALATLEAPARRLGNA--CAAAAPAAEAP	64			
Db	25	QPCGVILRIGSGSVRLGALPLPRAPRAARAARALAAALAPRLPHLSLELVAAAPAPDP	84			

QY	65	LSHWQALAEALAKPKDDPTFRNHCYQVA-PAVRVAIAETLGLACPAATTPKOLAEGRRLG	123
Db	85	ASLRGGLCOALVPP-----GVAALLFFPEARPELTLHFLAAATETP--VLSTLRR--	133
QY	124	FDEVFDTLFGADTLTMEEGSELLNRLTETHELEAHPHSDEPLRPFETSCCPNMIMLEKSY--	181
Db	134	--EARAFL-GAPLPHASPLETLLDYLVAVLQAHMEDVGLACIRPDQGGVALMTSRAG	190
QY	182	--PDLIPVYSCKSPOMMIAAMVYSYLAEEKKGIAPKADVMVMSIMPCTRKSEADRDWFCV	239
Db	191	RPQVLVDLSRDRDQDAGLRARLAPMAAPVGEAR-----	225
QY	240	DADPTLROLDHYITTVTELG-NIFKERGINLALPEGEENDPFGVSGAGVLFGTGGVME	298
Db	226	-----VPAAVLLGCDIARARRLVLEAVPPPHM-----TLGTT--LPP	260
QY	299	AALFRAYELFTGTPRLTSLSEVR-----GMDGIKENTITWVPAP-	338
Db	261	KALPTA-----GLRPGILLATGEVARNRPLEAIIHDIVQIVARALGSAAOVQPRALLPAV	315
QY	339	-----GSKFELLKRAAARAELAANHGP	362
Db	316	NCGLDLPACGSPGCFRLARFLANTSPFGRTGPRVWYTGSSQVNMSSHFKVMSLRDPRGAP	375
QY	363	GPLAWDGAGGETSEDR-----GGITLREVAANGLGNAKKL--ITKMOAGEAKYDFV---	412
Db	376	---AMATVGSW--RCGGDLEBPGGASARPPRPGAGQVMPKLRVYTLLE-----HPFVFR	425
QY	413	---ETMACPAC--CVGSGGGQPRSTDKAITQKQAAALYNLDEKSTLR-----	454
Db	426	DPDEGCGCPAQOLCL-----DPTNDSATLDALFALALANGSAPRALKCCGYCIDLLERTL	481
QY	455	SHENSTIRELY-----DYLGEPLGHKHAH	478
Db	482	AEDRPEDFELLYVGDKYCALRDGKMTGLVGDLLAGRAH	570

```

RESULT 5
US-09-934-070-10
; Sequence 10, Application US/09934070
; Publication NO. US20030092004A1
; GENERAL INFORMATION:
; APPLICANT: Lipton, Stuart A.
; APPLICANT: Zhang, Dongxian
; APPLICANT: Chatterton, Jon E.
; APPLICANT: Awobuluyi, Marc
; APPLICANT: Sevarino, Kevin A.
; TITLE OF INVENTION: EXCITATORY GLYCINE RECEPTORS AND METHODS
; FILE REFERENCE: P-LJ 4900
; CURRENT APPLICATION NUMBER: US/09/934,070
; CURRENT FILING DATE: 2001-08-20
; NUMBER OF SEQ ID NOS: 53
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10
; LENGTH: 897
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-934-070-10

Query Match          4.6%; Score 118; DB 9; Length 897;
Best Local Similarity 21.8%; Pred. No. 0.2;
Matches 128; Conservative 49; Mismatches 207; Indels 202; Gaps 25;

QY      7  KPCAAVSIRGSSCRAROVAPRAPLAASVYVALATLEAPARIRIGVNA--CAAAAPAAEAP 64
      :| | :| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB      25 QPCGYLALGGSVRLGALLPRAPPLARARAAALAAALAPRLPHNLSLELYVAAPARHP 84

QY      65 LSHVO---QAL-----AELAKPKDDP--TERHVCVQVAPVAAIAETGLAP-GATTP 112
      :| | :| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB      85 ASIRGCLQALVPEGVALLAFPEARPELQLHFLAAATETRPVLSILREARAPLGAANP 144

QY      113 KOLAEGLRRIGDEVEDTLTFGADLTIMEGSELLHRLTEHLAAPHSDPELPMTSCCPG 172

```

Db 145 FHL-----OLHNASPLET-----LLDVLVAVLQAHAMEDVGLACRTQDGC 185  
QY 173 KTLAMEKSY-----PDILPYVSSCKSPQMLAAVVKSYLAEKGIAPKDMVMSIMPCR 228  
Db 186 GLVALMTSTAGRPOLVLDLSRRDTGDAGLRARLAPMAAPVGGAP-----231  
QY 229 QSEADRWECVADPTLRQLDHYITTVTELG-NIFKRGINLAELPGEENDPWGSGAG 287  
Db 232 -----VPAVLLGCDIARARVLEAVPPCPHM-----258  
QY 288 VLEFGTGGVWEALRTAYELFTGTPLPRLSLSYR-----GMDGK 328  
Db 259 -LIGTP--LPPKALPTA-----GLPPGLALGEVAPRPLEAIIHVQVVARALGSAOV 310  
QY 329 ENTITWVAP-----GSKFEELKHRAAARAE 355  
Db 311 QPKRALLPAPVNCGLDQAPGESPGRFLANTSFQGTGPVWTGSSQVHMSRFKXWSLR 370  
QY 356 AAAGTTPGPLAMDGAGFTSEDR-----GGITLRVAVANGIGNAKRL--ITKMOAGEAK 408  
Db 371 RDRGAP---AMATVGSW--ROGQDLEPGASARPPPGAGVMPKRLRVTLLE-----420  
QY 409 YDFV-----EIMACFAG--CVGGGQPPSTDKAITOKROALYNLDEKSTLR-----454  
Db 421 HPEVFARDPDEDCQCPAGOLCT---DPGTNDSATIDALEFALANGSAPRALRKCXYC 476  
QY 455 -----SHENPSIRELY-----DYLGEPLGHKAH 478  
Db 477 IDLERLAEDTTPDFELIYVGDKYCALRDRGRTGLVGDLLAGRAH 522

## RESULT 6

US-09-934-070-6  
; Sequence 6, Application US/09934070  
; Publication NO. US2003092004A1  
; GENERAL INFORMATION:  
; APPLICANT: Lipton, Stuart A.  
; APPLICANT: Zhang, Dongxian  
; APPLICANT: Chatterton, Jon E.  
; APPLICANT: Awobuluyi, Marc  
; APPLICANT: Severino, Kevin A.  
; TITLE OF INVENTION: EXCITATORY GLYCINE RECEPTORS AND METHODS  
; FILE REFERENCE: P-LJ 4900  
; CURRENT APPLICATION NUMBER: US/09/934,070  
; CURRENT FILING DATE: 2001-08-20  
; NUMBER OF SEQ ID NOS: 53  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 6  
; LENGTH: 1011  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: VARIANT  
; LOCATION: 898, 899, 900, 901, 902, 903, 904, 905, 906, 907, 908, 909,  
; LOCATION: 910, 931, 932, 933, 934, 935, 936, 937, 938, 943, 944, 945,  
; LOCATION: 946, 947, 948, 949  
; OTHER INFORMATION: Xaa - Any Amino Acid  
US-09-934-070-6

Query Match 4.68; Score 118; DB 9; Length 1011;  
Best Local Similarity 21.88; Pred. No. 0.24;  
Matches 128; Conservative 49; Mismatches 207; Indels 202; Gaps 25;

QY 7 KPCAANSIRSSCAROVAPRAPLAASTVVALATLEAPARRIGNVA--CAAAPAAEAP 64  
Db 25 QPCGVLARIGSVRLGGLLEPRAPLARARALALARLRRLPHNLSLELVAAAPRAPDP 84  
QY 65 LSHVQ---QAL-----AELAKPRDDP--TRKHVCQVAVAVAAVLAETGLAP-GATTP 112  
Db 85 ASLIRGLCQALVPPGVAAALAFPEAPDELQHLFLAATETPVLSTLRBARAPLAPNP 144  
QY 113 KQLAEGRLRGTFDEVFTLTGADLTMEGSELLKRLTEHLEAHPHSDPELPMFTSCCPG 172

Db 145 FHL-----OLHNASPLET-----LLDVLVAVLQAHAMEDVGLACRTQDGC 185  
QY 173 KTLAMEKSY-----PDILPYVSSCKSPQMLAAVVKSYLAEKGIAPKDMVMSIMPCR 228  
Db 186 GLVALMTSTAGRPOLVLDLSRRDTGDAGLRARLAPMAAPVGGAP-----231  
QY 229 QSEADRWECVADPTLRQLDHYITTVTELG-NIFKRGINLAELPGEENDPWGSGAG 287  
Db 232 -----VPAVLLGCDIARARVLEAVPPCPHM-----258  
QY 288 VLEFGTGGVWEALRTAYELFTGTPLPRLSLSYR-----GMDGK 328  
Db 259 -LIGTP--LPPKALPTA-----GLPPGLALGEVAPRPLEAIIHVQVVARALGSAOV 310  
QY 329 ENTITWVAP-----GSKFEELKHRAAARAE 355  
Db 311 QPKRALLPAPVNCGLDQAPGESPGRFLANTSFQGTGPVWTGSSQVHMSRFKXWSLR 370  
QY 356 AAAGTTPGPLAMDGAGFTSEDR-----GGITLRVAVANGIGNAKRL--ITKMOAGEAK 408  
Db 371 RDRGAP---AMATVGSW--ROGQDLEPGASARPPPGAGVMPKRLRVTLLE-----420  
QY 409 YDFV-----EIMACFAG--CVGGGQPPSTDKAITOKROALYNLDEKSTLR-----454  
Db 421 HPEVFARDPDEDCQCPAGOLCT---DPGTNDSATIDALEFALANGSAPRALRKCXYC 476  
QY 455 -----SHENPSIRELY-----DYLGEPLGHKAH 478  
Db 477 IDLERLAEDTTPDFELIYVGDKYCALRDRGRTGLVGDLLAGRAH 522

## RESULT 7

US-09-737-149-33  
; Sequence 33, Application US/09737149  
; Patent No. US20020077466A1  
; GENERAL INFORMATION:  
; APPLICANT: Spaderna, Steven K  
; APPLICANT: Quinn, Kerry E.  
; APPLICANT: Shinkets, Richard A.  
; APPLICANT: Muralidhara, Padigaru  
; APPLICANT: Szytek, Kimberly A.  
; TITLE OF INVENTION: Polypeptides and Nucleic Acids Encoding Same  
; FILE REFERENCE: 15966-620 CIP  
; CURRENT APPLICATION NUMBER: US/09/737,149  
; CURRENT FILING DATE: 2001-06-15  
; PRIOR APPLICATION NUMBER: 60/170,564  
; PRIOR FILING DATE: 1999-12-14  
; PRIOR APPLICATION NUMBER: 60/173,165  
; PRIOR FILING DATE: 1999-12-27  
; PRIOR APPLICATION NUMBER: 60/173,362  
; PRIOR FILING DATE: 1999-12-27  
; PRIOR APPLICATION NUMBER: 60/173,544  
; PRIOR FILING DATE: 1999-12-29  
; PRIOR APPLICATION NUMBER: 60/174,404  
; PRIOR FILING DATE: 2000-01-04  
; PRIOR APPLICATION NUMBER: 60/174,962  
; PRIOR FILING DATE: 2000-01-07  
; PRIOR APPLICATION NUMBER: 60/223,929  
; PRIOR FILING DATE: 2000-08-09  
; NUMBER OF SEQ ID NOS: 49  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 33  
; LENGTH: 901  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-737-149-33

Query Match 4.58; Score 116; DB 10; Length 901;  
Best Local Similarity 21.78; Pred. No. 0.31;  
Matches 128; Conservative 49; Mismatches 207; Indels 206; Gaps 25;

QY 7 KPCAANSIRSSCAROVAPRAPLAASTVVALATLEAPARRIGNVA--CAAAPAAEAP 64

```

Db      25  QPCGVLAIRLGSGSVRLGALLPRAPLARARARALAPRLPHNLSTLELVAAAPPARDP 84
QY      65  LSHVQ---OAL-----AELAKPKDDE--TRKHVCYQVAVAVATAETGLAP--GATTP 112
      85  ASLTRGLCOLVPPGVAALLAPPEARPELLQLHFLAAATETPVLSLRREARAPLCAPNP 144
QY      113 KOLAEGLRRGDEVEDTLEFGADLTMEESSELLHRLTELEHNPMSDEPLPFMTSCCPG 172
      145 FHL-----QLHMASPLET-----LLDVLVAVLQAHAMEDVGLACRQDDPG 185
Db      173  WIMLEKSY---PDLIPYVSSCKSPOMLAAVVKSYLAEKGIAPKDMVMSIMPTCKR 228
      186  GLVALMTSRAGRPQVLYLDSRRDTGDAGLRARLAPMAAPVSGEAP----- 231
QY      229  QSEADRDWFCVDADPTLRQLDHYITTYVEIG--NIEKRGIMLAEIPGEMDNPMGVSGAG 287
      232  -----VPAAVLLGCDIARARVLEAVPCGPHM----- 258
QY      288  VLEGTGGVMEALRTAYELFTGTPLRLSLSEVR-----GMDGIR 328
      259  -LLGTP--LPPKALPTA-----GLPPGLALGEVAPRPLEAITHDYQVVARALGSAQV 310
Db      329  ETNITVVPAP-----GSKFEELKHRAA 351
      311  QPKRALLPAPVNGDQLQAPGESPGRFLAFLANTSPQRTGFWWTGSSQVHMSRHFV 370
QY      352  ARAEAAAHGTTPGLANDGAGFTSEDR-----GGITLRVAVANGIGNAKKL--ITRMOA 404
      371  WSLRRPGRAP---ANATVGSW--RDQQLDLLEGGSARPPPOGAQVWPKLRYVTILE- 424
QY      405  GEAKYDFV-----ETMACPAG--CVGGGGQPRSTDKATQKROALYNLDEKSTLR-- 454
      425  ---HPEVFARDDEDDGQCPAGQLCL---DPGTNDSATLDALFALANSGAPRALKKCC 476
QY      455  -----SHENPSIRELY-----DYLGEPLGKHAK 478
      477  YGYCIDLERLAEDTPDEFELLYVGDKYKALRDGRWTGLVGDLLAGRAH 526
Db

```

RESULT 8  
 US-09-737-149-8  
 ; Sequence 8, Application US/09737149  
 ; Patent No. US20020077466A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Spaderna, Steven K  
 ; APPLICANT: Quinn, Kerry E.  
 ; APPLICANT: Shimkets, Richard A.  
 ; APPLICANT: Muralidhara, Padigaru  
 ; APPLICANT: Szytek, Kimberly A.  
 ; TITLE OF INVENTION: Polypeptides and Nucleic Acids Encoding Same  
 ; FILE REFERENCE: 15966-620 CIP  
 ; CURRENT APPLICATION NUMBER: US/09/737,149  
 ; CURRENT FILING DATE: 2001-06-15  
 ; PRIOR APPLICATION NUMBER: 60/170,564  
 ; PRIOR FILING DATE: 1999-12-14  
 ; PRIOR APPLICATION NUMBER: 60/173,165  
 ; PRIOR FILING DATE: 1999-12-27  
 ; PRIOR APPLICATION NUMBER: 60/173,362  
 ; PRIOR FILING DATE: 1999-12-27  
 ; PRIOR APPLICATION NUMBER: 60/173,544  
 ; PRIOR FILING DATE: 1999-12-29  
 ; PRIOR APPLICATION NUMBER: 60/174,404  
 ; PRIOR FILING DATE: 2000-01-04  
 ; PRIOR APPLICATION NUMBER: 60/174,962  
 ; PRIOR FILING DATE: 2000-01-07  
 ; PRIOR APPLICATION NUMBER: 60/223,929  
 ; PRIOR FILING DATE: 2000-08-09  
 ; NUMBER OF SEQ ID NOS: 49  
 ; SOFTWARE: PatentIn Ver. 2.0  
 ; SEQ ID NO 8  
 ; LENGTH: 1043  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens

```

US-09-737-149-8
Query Match      4.5%; Score 116; DB 10; Length 1043;
Best Local Similarity 21.7%; Pred. No. 0.38;
Matches 128; Conservative 49; Mismatches 207; Indels 206; Gaps 25;

```

Query 7 KPCAAYSIRGSSCRARQVAPRAPLAATVRAATLAEAPARRIGVNA--CAAAAPAAEAP 64  
 Db 25 QPCGVLAIRLGSGSVRLGALLPRAPLARARARALAPRLPHNLSTLELVAAAPPARDP 84  
 QY 65 LSHVQ---OAL-----AELAKPKDDE--TRKHVCYQVAVAVATAETGLAP--GATTP 112  
 Db 85 ASLTRGLCOLVPPGVAALLAPPEARPELLQLHFLAAATETPVLSLRREARAPLCAPNP 144  
 QY 113 KOLAEGLRRGDEVEDTLEFGADLTMEESSELLHRLTELEHNPMSDEPLPFMTSCCPG 172  
 Db 145 FHL-----QLHMASPLET-----LLDVLVAVLQAHAMEDVGLACRQDDPG 185  
 QY 173 WIMLEKSY---PDLIPYVSSCKSPOMLAAVVKSYLAEKGIAPKDMVMSIMPTCKR 228  
 Db 186 GLVALMTSRAGRPQVLYLDSRRDTGDAGLRARLAPMAAPVSGEAP----- 231  
 QY 229 QSEADRDWFCVDADPTLRQLDHYITTYVEIG--NIEKRGIMLAEIPGEMDNPMGVSGAG 287  
 Db 232 -----VPAAVLLGCDIARARVLEAVPCGPHM----- 258  
 QY 288 VLEGTGGVMEALRTAYELFTGTPLRLSLSEVR-----GMDGIR 328  
 Db 259 -LLGTP--LPPKALPTA-----GLPPGLALGEVAPRPLEAITHDYQVVARALGSAQV 310  
 QY 329 ETNITVVPAP-----GSKFEELKHRAA 351  
 Db 311 QPKRALLPAPVNGDQLQAPGESPGRFLAFLANTSPQRTGFWWTGSSQVHMSRHFV 370  
 QY 352 ARAEAAAHGTTPGLANDGAGFTSEDR-----GGITLRVAVANGIGNAKKL--ITRMOA 404  
 Db 371 WSLRRPGRAP---ANATVGSW--RDQQLDLLEGGSARPPPOGAQVWPKLRYVTILE- 424  
 QY 405 GEAKYDFV-----ETMACPAG--CVGGGGQPRSTDKATQKROALYNLDEKSTLR-- 454  
 Db 425 ---HPEVFARDDEDDGQCPAGQLCL---DPGTNDSATLDALFALANSGAPRALKKCC 476  
 QY 455 -----SHENPSIRELY-----DYLGEPLGKHAK 478  
 Db 477 YGYCIDLERLAEDTPDEFELLYVGDKYKALRDGRWTGLVGDLLAGRAH 526

RESULT 9  
 US-09-976-059-15  
 ; Sequence 15, Application US/09976059  
 ; Patent No. US20020164747A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Farnet, Chris  
 ; APPLICANT: Zazopoulos, Emmanuel  
 ; APPLICANT: Staifa, Alfredo  
 ; TITLE OF INVENTION: Genes and Proteins for Biosynthesis of Ramoplanin  
 ; FILE REFERENCE: 3019-PCT  
 ; CURRENT APPLICATION NUMBER: US/09/976,059  
 ; CURRENT FILING DATE: 2001-10-15  
 ; NUMBER OF SEQ ID NOS: 34  
 ; SOFTWARE: PatentIn version 3.0  
 ; SEQ ID NO 15  
 ; LENGTH: 4999  
 ; TYPE: PRT  
 ; ORGANISM: Actinoplanes sp.  
 US-09-976-059-15

Query Match 4.5%; Score 116; DB 9; Length 4999;  
 Best Local Similarity 24.1%; Pred. No. 3.5; Indels 112; Gaps 20;  
 Matches 100; Conservative 32; Mismatches 171;

Query 8 PCAAVSIRGSSCRARQVAPRAPLAATVRAATLAEAPARRIGVNAACAAAPAAEAP 64  
 Db 25 QPCGVLAIRLGSGSVRLGALLPRAPLARARARALAPRLPHNLSTLELVAAAPPARDP 84

Db 4400 PDSASQVAMHR-RQLAGAPDELPLADHPRPAEATYR-----GHTVEFVPA----- 4448  
QY 65 LSHVQALALAPLAKDDPTKKHVCYAPAVVAIAETLG-LAPGATTPQLAEGRLG 123  
Db 4449 ---VHQALAEIARRG-----VYFMTVQALANLLSKLAGADIPIGVAVAGRT-- 4495  
QY 124 FDEVDPLFG-----ADLTMEGSELLHRLTE-HLEAHPSDEPLPMFTSCPGW 173  
Db 4496 -DPTLDLIGFVNTLVLRDLDGNGPTTDLHRTTRTTLHAFTHODVP----- 4543  
QY 174 IAMEKSYDPLIPYVSSCKSP--OMMLAANVKSYLEKKCIAPKDMVMSIMPCRKQSE 221  
Db 4544 ---FEKLVEDLAPTRSLARHPLFQVMTTQSTGRAGAAELPGLEFVILSPGVAAK--- 4597  
QY 232 ADGMFCVADPTLRQDHNVTTELGNIFKERGINALPEGBEMDPMGVSGAGVLF 291  
Db 4598 -----VDLD-----LSISEAVDDG-----RPAQL---AGLVLA 4623  
QY 292 TTG---GVMEALRTAYELFTGPTL-PRLSLSEVRGMDG-----IKETNITWVPAGS 340  
Db 4624 AADLFEGHTAERIAGYLAIRLAVLPADPGARLDVDLDLGEERRIVLTGMNDTTAANPAV 4683  
QY 341 KPEBLKRRMAARAARAAHGTPOGLANDGA--GFTSDEGRGGITLRVAVANGLG 393  
Db 4684 AVBELIERRAAAEPEAGA-----WVCGDTHLRYGELNARNLALFLVERGAG 4731

## RESULT 10

US-10-047-542-75  
; Sequence 75, Application US/10047542  
; Patent No. US20020168367A1  
; GENERAL INFORMATION:  
; APPLICANT: LARIBICK, JAMES W.  
; APPLICANT: KYCOFF, KEITH L.  
; TITLE OF INVENTION: NOVEL IMMUNOADHESINS FOR TREATING AND PREVENTING VIRAL  
; TITLE OF INVENTION: AND BACTERIAL DISEASES  
; FILE REFERENCE: 030905.0004.CIP1  
; CURRENT APPLICATION NUMBER: US/10/047,542  
; PRIOR FILING DATE: 2001-10-26  
; PRIOR APPLICATION NUMBER: PCT/US01/13932  
; PRIOR FILING DATE: 2001-04-28  
; PRIOR APPLICATION NUMBER: 60/200,298  
; PRIOR FILING DATE: 2000-04-28  
; NUMBER OF SEQ ID NOS: 101  
; SOFTWARE: Patent In Ver. 2.1  
; SEQ ID NO 75  
; LENGTH: 912  
; TYPE: PRT  
; ORGANISM: Oryctolagus cuniculus  
US-10-047-542-75

Query Match 4.3%: Score 111; DB 9; Length 912;

Best Local Similarity 20.9%; Pred. No. 0.84; Indels 166; Gaps 26;

Matches 114; Conservative 56; Mismatches 210;

QY 18 SCARQVAPRAPLAASTVVALATLE-----APARRIGNVCAAAAPAAE-----A 63  
Db 138 SCARPGAGPGRSLTLTRGAGDELIRRSFAGEPARAGAVLTAVLARREDHGANFSCRA 197  
QY 64 PLSHVQALAE-----ELAKPKDDPTKKHVCYAPAVVAIAETL-----GL 105  
Db 198 ELDRPQGLALFENSSAPROLMTYALPLDSPR-----LLAPRVLEVDQSLSVCTLDGL 251  
QY 106 APGATTPQLAEGRLGFEVDPLFG-----ADLTMEGSELLHRLT---EHLA 155  
Db 252 PPAEAGVHIALGKRLN-PEV--TLEGDAIVATATTAEEGKQLVCAVTLGGERRES 308  
QY 156 HPH---SDEPLPMFTSCPGWIMALEKSYDPLIPYVSSCKSPOMMLAANVKSYLEKKGI 212  
Db 309 RENTVVSFAPRLT-----LSEPSAPGKLVYVTCAGARALVTL-----EGV 352  
QY 213 APKDMVMSIMPCRKQSEADRMFCVADPTLRQDHNVT---TVELGNIFKERGINLA 269

Db 353 PAAPAGPQALQFNASESDDGSRFFC---DATLELDJETLSKNGSAELRVLYAPR-LDDA 408  
QY 270 ELPEGEENDPMG-----VSGAGVLFETTTGGVMEALRTAYELF 308  
Db 409 DCPR-SWTWPEGEQGLRCEARGNPPPAVHCAASDGAVALALGLIPVTRALAGTY---- 463  
QY 309 TGTPPLRLSLSEYRGMDGIKETNITWVPAP-----GSKFEELKRRMAARAARAAAGTPG 363  
Db 464 -----RCTAANVG--EAVNDVYLTVEYAPALDSVGCPEKVTYLTGTEASLSCVAGVPR 516  
QY 364 PLAMDGAGFTSDEGRGT-----TLRVAVANGLSNAKLLITRMQAGEAKTYDEVI 414  
Db 517 PSV--SCVRROADVIEGLLVAREHAGTYRCEAINARALAKVAVATVEYGPS----FEE 570  
QY 415 MACPAG-----CVGGG-----OPRSTDKAITOK 438  
Db 571 RSCPSTMTWVGSSEQLFSCVEBGRPOPSVOCVSGASEGLPLAPLAPNPSDPSPVRD 630  
QY 439 ROAALY 444  
Db 631 LARGIV 636

## RESULT 11

US-09-815-242-10258  
; Sequence 10258, Application US/09815242  
; Patent No. US20020061569A1  
; GENERAL INFORMATION:  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Karl L.  
; APPLICANT: Zyskind, Judith W.  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John D.  
; APPLICANT: Carr, Grant J.  
; APPLICANT: Yamamoto, Robert T.  
; APPLICANT: Xu, H. Howard  
; TITLE OF INVENTION: Identification of Essential Genes in  
; TITLE OF INVENTION: Prokaryotes  
; FILE REFERENCE: ELITRA.011A  
; CURRENT APPLICATION NUMBER: US/09/815,242  
; PRIOR FILING DATE: 2001-03-21  
; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/242,578  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: 60/253,625  
; PRIOR FILING DATE: 2000-11-27  
; PRIOR APPLICATION NUMBER: 60/257,931  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/269,308  
; PRIOR FILING DATE: 2001-02-16  
; NUMBER OF SEQ ID NOS: 14110  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 10258  
; LENGTH: 857  
; TYPE: PRT  
; ORGANISM: Escherichia coli  
US-09-815-242-10258

Query Match 4.3%: Score 109.5; DB 10; Length 857;

Best Local Similarity 21.2%; Pred. No. 1; Indels 91; Gaps 19;

Matches 86; Conservative 67; Mismatches 162;

QY 108 GATTPQLAEGRLR-LGFDEVDPLFGADLTMEGSELLHRLTENLEAHP---SDEPL 163  
Db 312 GATLDDRYRIENDALERFQKVFVAEDSV-EDTIALIRGLKEVELNHNQITPRAI 370  
QY 164 PMFTSCPGWIMALEKSYDPLIPYVSSCKSPOMMLAANVKSYLEKKGI 199

Db 371 VAAATLSHRYIA--DRQLPDKAIDLIDEAASSIRMOIDSKPEELRLDRRIITOLKLEQQA 428  
OY 200 AMVKSYLAERKGIAPDMVMSIMPCTRKSEADRMFCVAD---PTLR-QLDHVIT 254  
Db 429 LMKEDDEASKRL---DMINELSDKERYISELEEMKAEKASLSGTOTIKALEEAKTA 485  
OY 255 VE---LGNIFKFERGINIAELPEGEWDMVMGSGAVLEFGTT---CGVMEALRTY 305  
Db 486 IQOARVGDALMSELSQYKIPLEKQL-----EATQLEGKTMRLRNKVTDAELAEVL 540  
OY 306 ELFTGTPLRLSLSEYRKMDGKIKETNITVWPAPGSEFELLKRAAAREAA----- 356  
Db 541 ARMTGIPVSRMSEB-----REKLIRMA-----EQLHHRVIGQNEAVDAVSNAIR 585  
OY 357 -AAHGTGGLAMDGAGAGTSEDGRCITLRVANGLAKKLITKMGAE--AKYDFVE 413  
Db 586 RSRAGLADPNRPIGSLFLGPTGVGKTELCKALANMFSDAMVRIDSEMEKHSVR 645  
OY 414 IMACPAGCVG---GGGQPRSTDKAITOKROALYNLDEKSTLRSH 456  
Db 646 LVGAPPGVYGEYEGGVLTEA-----VRRRPYSVITLDE---VEKMH 683

## RESULT 12

US-09-934-070-8  
; Sequence 8, Application US/09934070  
; Publication No. US20030092004A1  
; GENERAL INFORMATION:  
; APPLICANT: Lipson, Stuart A.  
; APPLICANT: Zhang, Dongxian  
; APPLICANT: Chatterton, Jon E.  
; APPLICANT: Awobuluyi, Marc  
; APPLICANT: Sevarino, Kevin A.  
; TITLE OF INVENTION: EXCITATORY GLYCINE RECEPTORS AND METHODS  
; FILE REFERENCE: P-12 4900  
; CURRENT APPLICATION NUMBER: US/09/934,070  
; CURRENT FILING DATE: 2001-08-20  
; NUMBER OF SEQ ID NOS: 53  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 8  
; LENGTH: 1005  
; TYPE: PRT  
; ORGANISM: Mus musculus  
US-09-934-070-8

Query Match 4.2%; Score 108; DB 9; Length 1005;

Best Local Similarity 22.1%; Pred. No. 1.7;

Matches 121; Conservative 52; Mismatches 177; Indels 198; Gaps 28;

OY 7 KCAAASIRSSCORANQVAPRAPIAATVRAVALATEAPARLGNVACAAAPAAAPIS 66  
Db 27 QCGCVTRAGASVRLAALPRAPARARVLAALAT---PSPRLPH----- 68  
OY 67 HVOQALAEALAKRPDPT--RKHYCVQVAPAVRAVIAETGLAPGATTPQLAEGRLRIG 124  
Db 69 NLSLELVANASPTROPASLARGLCQVLA-----PGVV----- 101  
OY 125 DEVFDTLFGADLTIMESSELRLHRLTEHLEAHHSDEPL-----PMFTSCPP--- 171  
Db 102 -----ASITTFEAREPL--RLQFLAA--ATETPVLSVLRREVRAPLGRATPFHL 148  
OY 172 --GMIALEKSYPDLLPYVSSCKSPOMMLAAVKSYLEAKGIAPDMVMSIMPCTRKQ 229  
Db 149 QIDMASPLETIDVILVSLVRAHAMEDI---ALVLCVRRPSGL-----VLTMTSR 196  
OY 230 SEADRMFCVADPTLRQLDHVTITVELGNIFKERCINIAELPEGEWDMVMGSGAGVL 289  
Db 197 SQAPK--FVID-----LSQDS-----GNDSLRATLALGLTEGG---GTPVSAAYL 238  
OY 290 FGTGGS---VMEALRTYELFTGTPRLSLSEV---RQMDGKIKETNITVWPAPGSKF 342  
Db 239 LQSTHAHEVLEAA--PPSPOMLGTPLPAEALPRTGLPPGVLVLEGTGCPISLEAAVHDM 297

OY 343 EEL-----LKHRAAAREAA-----AHGTPGPLA-----W 367  
Db 298 VELVRAALSSMALMHERALPLPAVNCEDLKTGSSSESTARTLARMLNSTSPQGRGAYR 357  
OY 368 DQAGAGTS-----EDGR-----GCTTLVAVANGIGNAK 396  
Db 358 VAGSSQVHVSRRFKWLSLRDPLGAPAMATVGSMDGQIDFQGAALRVPSPSGTQARP 417  
OY 397 KL--ITKMGAGEAKYDFV-----EIMACPAG--CVGGGGGQPRSTDKAITOKROALYNL 446  
Db 418 KLRVTLVE---HPFVETRSDSDGQCPAQQLC---DPGTNDSARLDALFTALENG 468  
OY 447 DEKSTLR 454  
Db 469 SVPRITLR 476

## RESULT 13

US-10-188-246-12  
; Sequence 12, Application US/10188246  
; Publication No. US20030087274A1  
; GENERAL INFORMATION:  
; APPLICANT: Anderson, David W.  
; APPLICANT: Boldog, Ferenc,  
; APPLICANT: Casman, Stacie,  
; APPLICANT: Edinger, Shlomil,  
; APPLICANT: Gerlach, Valerie,  
; APPLICANT: Gorman, Linda,  
; APPLICANT: Li, Li,  
; APPLICANT: Malvankar, Uriel,  
; APPLICANT: Patursajan, Meera,  
; APPLICANT: Peyman, John,  
; APPLICANT: Shenoy, Suresh,  
; APPLICANT: Shimkets, Richard,  
; APPLICANT: Vernet, Corine A. M.,  
; APPLICANT: Voss, Edward  
; TITLE OF INVENTION: THERAPEUTIC POLYPEPTIDES, NUCLEIC ACIDS ENCODING SAME, AND MET  
; FILE REFERENCE: 21402-397B US  
; CURRENT APPLICATION NUMBER: US/10/188,246  
; CURRENT FILING DATE: 2002-07-02  
; PRIOR APPLICATION NUMBER: 60/303046  
; PRIOR FILING DATE: 2001-07-05  
; PRIOR APPLICATION NUMBER: 60/304502  
; PRIOR FILING DATE: 2001-07-11  
; PRIOR APPLICATION NUMBER: 60/305011  
; PRIOR FILING DATE: 2001-07-12  
; PRIOR APPLICATION NUMBER: 60/306085  
; PRIOR FILING DATE: 2001-07-17  
; PRIOR APPLICATION NUMBER: 60/326981  
; PRIOR FILING DATE: 2001-10-14  
; PRIOR APPLICATION NUMBER: 60/360923  
; PRIOR FILING DATE: 2002-02-28  
; PRIOR APPLICATION NUMBER: 60/363636  
; PRIOR FILING DATE: 2002-03-12  
; PRIOR APPLICATION NUMBER: 60/373063  
; PRIOR FILING DATE: 2002-04-16  
; NUMBER OF SEQ ID NOS: 60  
; SOFTWARE: Custom  
; SEQ ID NO 12  
; LENGTH: 450  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-188-246-12

Query Match 4.2%; Score 107.5; DB 9; Length 450;

Best Local Similarity 20.5%; Pred. No. 0.61;

Matches 92; Conservative 53; Mismatches 152; Indels 151; Gaps 20;

OY 17 SSCRAQVAPRAL--AASTVRAVALA--TLAP-----ARPLGVACAAAPAAAPAEPL 65  
Db 15 SACR-----PRDLLEAAVYLRAGAGPVSPGGGGGGGGGRTLAQAAAGAAVPAAYPR 69  
OY 66 SHVOALAEALAKRPDPTTRKHCVCVY-----APAVRAVAI-----AETL--GLAPGAT 110

```

Db 70 ARAARRAGSGFRNGSVPRHFMMSLYSLAGRAPGAAGAAVSAGHGADITINGFTDQAT 129
Qy 111 TPQOLABGLRRLCDP-----EVTFTLGADITIMEBSELHRLTHELHAHPHS--DEPLPM 165
Db 130 QDSAAETGOSFLFEDVSLINDADEVVGAEFLRYLRGSP-----ESGPGSMTSPPL 180
Qy 166 FTSCCGMTAMEKSPDLIPYVSSCKSPQMLAAVKSYLEAKKGIAPKDMVMSIMPC 225
Db 181 LSTCGG-----AARAPRLYSRAAEPLVGQRW-----EAFVADAMRR 219
Qy 226 TRKQSEADRMFCVADPTLRQLDHYITVELGNIFKRGINLAELPEGEMDPKMGVSG 285
Db 220 HREPRPPRA-FCL-----LLRAVAGVPSPPLV--- 247
Qy 286 AGVLEFGTGVMEALRTAELFTGTPRLSLSEVRGMGDIKETNITWYPAGSKFEEL 345
Db 248 RRLGFGMPGG-----GGSAEERAVLVSSRTQRESL 280
Qy 346 LKH-RAAARAEAAA-----HGTGPELWMDG-----AGTSEDRGGITLRY 386
Db 281 FREIRAOARALGAALASEPLDPDGTGASPRAYIGRRRRRTALAGTRTQSGSG----- 335
Qy 387 AVANGMAKKLITKMOAGEAKYDFEI 414
Db 336 --GAGRGHGRGRSRSRKPLHYDFEKL 361

```

```

RESULT 14
US-09-934-070-2
; Sequence 2, Application US/09934070
; Publication No. US20030092004A1
; GENERAL INFORMATION:
; APPLICANT: Lipton, Stuart A.
; APPLICANT: Zhang, Dongxiam
; APPLICANT: Chatterton, Jon E.
; APPLICANT: Awobuluyi, Marc
; APPLICANT: Severino, Kevin A.
; TITLE OF INVENTION: EXCITATORY GLYCINE RECEPTORS AND METHODS
; FILE REFERENCE: P-LJ 4900
; CURRENT APPLICATION NUMBER: US/09/934,070
; CURRENT FILING DATE: 2001-08-20
; NUMBER OF SEQ ID NOS: 53
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 987
; TYPE: PRT
; ORGANISM: Rattus sp.
US-09-934-070-2

Query Match 4.28; Score 107.5; DB 9; Length 987;
Best Local Similarity 22.08; Pred. No. 1.9;
Matches 114; Conservative 48; Mismatches 200; Indels 157; Gaps 23;

Qy 7 KPCAASIRGSSCRAROVAPRAPLAASTVVALATLEAPARRLGNVCAAAAPAAEAPLS 66
Db 27 QPCRVPTRAGASVRLAALPRAPARARVLAALAT---PAPRLPH----- 68
Qy 67 HVOQALAEKAPKDDPT--RKHYCVQVAPRAVAIAETLGLAGATTPKOLABGLRLGF 124
Db 69 NLSLELVAVASPTRPDASLARGLCOVLAAP-----GVVASIAPPEARPE-LRLLOF 118
Qy 125 ----DEVFDTLFGADLTMEGSELHRLTHELHAHPHSDEPLPMFTSCPGMTAMEKS 180
Db 119 LAATETPTVPRFHQLDMASPLETILDVLSVRAHAMEDIALYLCAVVRPGSLVTLMTN 178
Qy 181 Y----PDLIPYVSSCKSPQMLAAVKSYLEAKKGIAPKDMVMSIMPCRKQSEADRMV 236
Db 179 HASGAPKFVLDLSRDSLRNSLRAGLALLGALGGGTPVPAAV--LLGSGTARANE---- 232
Qy 237 FCVDADPTLRQLDHYITVELGNIFKRGINLAELPEGEM--DNPM-----GVSGA 286
Db 233 -VLEAAP-----PGQMLDGTPLPAEALPTTGILPPGV 263

```

```

Qy 287 GVLFTTGGVMEALRTAELF-----TGTP 313
Db 264 LALGETEHSLEAVDHWELVAQALSSNALVHPERALLPAVNCDDLTGSGSEATGR 323
Qy 314 PR-LSLSEVRGMGDIKETNITWYPAGSKFEELKHAARAABAAHGTGPPPLAW----- 367
Db 324 ARFLGNTSFGGRG-----AWVTGSSQVHVSRIKFWMSLRDPLGAP--AMATVGS 373
Qy 368 --DGGAGFTSEDRGGITLRVAVANGLNAAKRL--ITKMOAGEAKYDFV-----ETMNC 417
Db 374 WQDQGLDFQP-----GAALRVPSGTOARPKLRVTVLVE-----HPFVFTRESDEGOC 424
Qy 418 PAG--CVGGGCGPRSTDKATQRQAALYLNDEKSTLRR 454
Db 425 PAGQLCL-----DQGTNDSARLDLFPALVAVGSPRLRR 459

```

```

RESULT 15
US-09-934-070-4
; Sequence 4, Application US/09934070
; Publication No. US20030092004A1
; GENERAL INFORMATION:
; APPLICANT: Lipton, Stuart A.
; APPLICANT: Zhang, Dongxiam
; APPLICANT: Chatterton, Jon E.
; APPLICANT: Awobuluyi, Marc
; APPLICANT: Severino, Kevin A.
; TITLE OF INVENTION: EXCITATORY GLYCINE RECEPTORS AND METHODS
; FILE REFERENCE: P-LJ 4900
; CURRENT APPLICATION NUMBER: US/09/934,070
; CURRENT FILING DATE: 2001-08-20
; NUMBER OF SEQ ID NOS: 53
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 1002
; TYPE: PRT
; ORGANISM: Rattus sp.
US-09-934-070-4

Query Match 4.08; Score 104; DB 9; Length 1002;
Best Local Similarity 22.18; Pred. No. 3.8;
Matches 117; Conservative 48; Mismatches 201; Indels 164; Gaps 25;

Qy 7 KPCAASIRGSSCRAROVAPRAPLAASTVVALATLEAPARRLGNVCAAAAPAAEAPLS 66
Db 27 QPCRVPTRAGASVRLAALPRAPARARVLAALAT---PAPRLPH----- 68
Qy 67 HVOQALAEKAPKDDPT--RKHYCVQVAP--AVRAVAIETLC-----LAPGATTPO 114
Db 69 NLSLELVAVASPTRPDASLARGLCOVLAAPGVVASIAPPEARBLRLLOFLAATETP-- 126
Qy 115 LAGLIR-----LGFVDFPTLTGADLTMEBSELHRLTHELHAHPHSDEPLPMTSC 169
Db 127 VVSLRREVRTALG---ADTPHQLQDMASPLETILDVLSVRAHAMEDIALVLCRR 182
Qy 170 CPGMIMALEKSY----PDLIPYVSSCKSPQMLAAVKSYLEAKKGIAPKDMVMSIMPC 225
Db 183 DPGSLVLTMTNHAQAPKFVLDLSRDSLRNSLRAGLALLGALGGGTPVPAAV--LLGC 240
Qy 226 TRKQSEADRMFCVADPTLRQLDHYITVELGNIFKRGINLAELPEGEM--DNPM--- 280
Db 241 STARANE-----VLEAAP-----PGQMLDGTPLPAE 267
Qy 281 ----GVSGAGVLEFGTGVMEALRTAELF----- 308
Db 268 ALPTTGILPPVLALETGEHSLAVDHWELVAQALSSNALVHPERALLPAVNCDDLK 327
Qy 309 -----TGTPLR-LSLSEVRGMGDIKETNITWYPAGSKFEELKHAARAABAAHGT 361
Db 328 TGSGEATGRILARFLGNTSFGGRG-----AWVTGSSQVHVSRIKFWMSLRDPLGA 380
Qy 362 PGPLAW-----DGGAGFTSEDRGGITLRVAVANGLNAAKRL--ITKMOAGEAKYDFV 412

```

Db 381 P---AMATVGSWQDQLDFQP---GAALRVPSPSGTOARRPRLRVTLVE-----HPV 428  
QY 413 -----EIMACPAG---CYGGGGQPRSTDKAITOKROALYNIDKSTLRR 454  
Db 429 FTREDEDEGQCPAGQCL---DPTNDSARLDALFAALVNGSVPTLRR 474

Search completed: June 3, 2003, 16:17:40  
Job time : 26 secs

**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run On: June 3, 2003, 15:34:37 ; Search time 72 Seconds

(without alignments)  
919.799 Million cell updates/sec

Title: CAC80065

Perfect score: 2576  
Sequence: 1 MSALVKPCAAVSIKSSCR.....HELLHTHVAGVEKDEKK 497

Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database :

A.Geneseq\_101002:\*  
1: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:\*  
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:\*  
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:\*  
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:\*  
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:\*  
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:\*  
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:\*  
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:\*  
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:\*  
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:\*  
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:\*  
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:\*  
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:\*  
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:\*  
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:\*  
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:\*  
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:\*  
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:\*  
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:\*  
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:\*  
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:\*  
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:\*  
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	456.5	17.7	476	22	Human oxidoreducta
2	361.5	14.0	456	22	Human NADP hydroge
3	361.5	14.0	456	22	Human protein sequ
4	211.5	8.2	241	22	Drosophila melanog
5	144	5.6	4999	23	Ramoplanin biosynt
6	125.5	4.9	748	20	Granulocytic Ehrli
7	125.5	4.9	748	21	Granulocytic Ehrli
8	125.5	4.9	748	21	Granulocytic Ehrli
9	125.5	4.9	748	21	Granulocytic Ehrli
10	122.5	4.8	573	20	Ehrlichia sp. E82.

11	122	4.7	965	22	AAU02197
12	122	4.7	971	22	AAU02198
13	116	4.5	901	23	AAO18086
14	116	4.5	1043	22	AAU02199
15	116	4.5	4999	23	AAO22159
16	114.5	4.4	11096	22	AAE10129
17	114	4.4	912	16	AAW75206
18	112	4.3	4472	18	AAW25260
19	110.5	4.3	451	21	AAV74914
20	110	4.3	577	17	AAW06554
21	109.5	4.3	857	20	AAW97702
22	109.5	4.3	857	22	AAU34665
23	109.5	4.3	4572	19	AAW52845
24	108.5	4.2	1887	23	ABO08801
25	108.5	4.2	1887	23	ABO08802
26	108.5	4.2	1887	23	ABO08803
27	108.5	4.2	1887	23	ABO08804
28	108.5	4.2	1887	23	ABO08805
29	108.5	4.2	1887	23	ABO08806
30	108.5	4.2	1887	23	ABO08807
31	107.5	4.2	447	21	AAV74915
32	107.5	4.2	3398	14	AAW44430
33	107	4.2	901	21	AAW42494
34	105.5	4.1	6797	22	AAW31558
35	105	4.1	284	22	ABG17245
36	104	4.0	421	21	AAW35747
37	104	4.0	2675	21	AAW07564
38	104	4.0	5532	21	AAW23752
39	104	4.0	5532	22	AAW65267
40	103.5	4.0	1799	23	AAW50359
41	103	4.0	401	21	AAW33383
42	102.5	4.0	451	21	AAV74916
43	102	4.0	722	21	AAV78916
44	102	4.0	739	21	AAV78926
45	102	4.0	756	21	AAV78927

#### ALIGNMENTS

RESULT 1  
ID: AAB73690 standard; Protein; 476 AA.  
AC: AAB73690;  
DE: 11-SEP-2001 (first entry)  
XX: Human oxidoreductase protein ORF-23.  
XX: Human oxidoreductase protein ORF-23.  
KW: Human oxidoreductase protein; ORF; cell proliferative disorder;  
KW: arteriosclerosis; cirrhosis; psoriasis; cancer; endocrine disorder;  
KW: diabetes mellitus; diabetes insipidus; dwarfism; hirsutism; amenorrhea;  
KW: osteoporosis; metabolic disorder; obesity; phenylketonuria;  
KW: hypercholesterolaemia; reproductive disorder; infertility;  
KW: ovulatory defect; menstrual cycle defect; endometriosis;  
KW: polycystic ovary disease; spermatogenesis disruption; impotence;  
KW: neurological disorder; epilepsy; stroke; Alzheimer's disease;  
KW: Huntington's disease; Parkinson's disease; Creutzfeldt-Jakob disease;  
KW: meningitis; cerebral palsy; muscular dystrophy; mood disorder; anxiety;  
KW: schizophrenic disorder; infection; autoimmune disorder;  
KW: inflammatory disorder; acquired immunodeficiency syndrome; AIDS; asthma;  
KW: allergy; Crohn's disease; atrophic dermatitis; gout; multiple sclerosis;  
KW: rheumatoid arthritis; ulcerative colitis; drug screening;  
KW: toxicity screening; transgenic animal; SNP detection; gene therapy.  
OS: Homo sapiens.  
PN: WO200144448-A2.  
PD: 21-JUN-2001.  
PF: 07-DEC-2000; 2000WO-US33158.



Oy	167	ISCCGGTATAMLEKSV--PDILIPVSSCKSPQNMHLAAVKSYLEAKGKIAPDMYVSGIMPC	225
Oy	167	: : :       : : :       : : :       : : :       : : :	
Dd	169	TSACDGNVRYAERVLGRPTTHHLCITAKSPQYMGSLVKOTFAHQNLSPKITHVYAPC	228
Oy	226	TRKQSEADRDWFCVADPTL---RQLDHYITTVELGNIKERGINLAELPEGEMDNPMGV	282
Dd	229	YDKRKLEALQE-----SLPALHSGRCADVLTSGEINQIMEQGLSVRD-----	272
Oy	283	GSGAGVLTG-----TTGGVMAALLRTAYELFTGTPPLRLSLSEVRGMDGIC	328
Dd	273	-AAVDTLTGDLKEDKVTTRHDGASSDGHILAHIFRFAKLEFN-----EDVE	316
Oy	329	ENITMVPAPGSKFEELKHKRAAAREAAAHQTPGLANDGCGFTSEDRGSGITLRVAV	388
Dd	317	E--VYRRALRNKDFCEVTLEK-----NGSVLRFAA	345
Oy	389	ANGICNAKLLITKMOAGEAKYDFVEIMACPGCVGGGQPRS---TDKAITOKROALY	444
Dd	346	AYGFENIQMILKLKKKPFHFVEVLACAGGCINCRGQAQPTDGNADKRLLRQMEIYA	405
Oy	445	NLDEKSTLRSHENPISIRELYDTYVLEPILGHKAHELLHNY	485
Dd	406	DI---PVRRPRESSAHVQELVQEWLEGINSPKAREVILHTTY	442
RESULT 3			
AAB92498			
ID	AAB92498	standard; Protein; 456 AA.	
AC	AAB92498;		
XX			
DT	26-JUN-2001	(first entry)	
XX			
DE	Human protein sequence SEQ ID NO:10602.		
XX			
KW	Human; primer; detection; diagnosis; antisense therapy; gene therapy.		
XX			
OS	Homo sapiens.		
XX			
PN	EP1074617-A2.		
XX			
PD	07-FEB-2001.		
XX			
PF	28-JUL-2000; 2000EP-0116126.		
XX			
PR	29-JUL-1999; 99JP-0248036.		
XX	27-AUG-1999; 99JP-0300253.		
PR	11-JAN-2000; 2000JP-0118776.		
XX	02-MAY-2000; 2000JP-0183767.		
PR	09-JUN-2000; 2000JP-0241899.		
XX			
PA	(HELI-) HELIX RES INST.		
XX			
PI	Ota T, Iisoga T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;		
PI	Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;		
XX			
DR	WPI: 2001-318749/34.		
XX			
PT	Primer sets for synthesizing polynucleotides, particularly the 5602		
PT	full-length cDNAs defined in the specification, and for the detection		
PT	and/or diagnosis of the abnormality of the proteins encoded by the		
XX	full-length cDNAs -		
PS	Claim 8; SEQ ID 10602; 2537pp + CD ROM; English.		
XX			
CC	The present invention describes primer sets for synthesizing 5602		
CC	full-length cDNAs defined in the specification. Where a primer set		
CC	comprises: (a) an oligo-dT primer and an oligonucleotide complementary		
CC	to the complementary strand of a polynucleotide which comprises one of		
CC	the 5602 nucleotide sequences defined in the specification, where the		
CC	oligonucleotide comprises at least 15 nucleotides; or (b) a combination		
CC	of an oligonucleotide comprising a sequence complementary to the		
CC	complementary strand of a polynucleotide which comprises a 5'-end		

[illegible]

PF 23-MAR-2001; 2001MO-US09231.  
XX  
XX 23-MAR-2000; 2000US-191637P.  
PR 11-JUL-2000; 2000US-0614150.  
XX  
XX (PEKE ) PE CORP NY.  
PI Venter JC, Adams M, Li PWD, Myers EW;  
XX WPI; 2001-656860/75.  
DR N-PSDB; ABL13486.  
XX  
XX New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signalling and cell-cell  
PT interactions -  
XX  
XX Disclousure; SEQ ID NO 34941; 21pp + Sequence Listing; English.  
XX  
XX The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from Drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
CC sequences (ABL01840-ABL16175) and the encoded proteins  
CC (AB57737-AB572072).  
CC The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
XX Sequence 241 AA;  
SQ  
Query Match 8.2%; Score 211.5; DB 22; Length 241;  
Best Local Similarity 24.6%; Pred. NO. 2e-12; Indels 67; Gaps 10;  
Matches 70; Conservative 47; Mismatches 101;  
OY 198 LAANVSKYLAEKGIAPKDMVWVSIMPCTKROSEADRDWCVDADPTLRQLDHYITVEL 257  
DB 1 MGVLYKQILDKNMVPASRIHYTVMPICYDKLEASRDEFSKANNNS-RDVDCVITVEV 59  
OY 258 GNIRKGINLAELP-----EGEDN---PMGVSGAGVLTGTTGGVNEALR-TAYELFT 309  
DB 60 EQLLESAOQPLSQVDLLDLPMSNVPEFVWVAHEKTL--SGVGEHIFKVAKKHIFN 116  
OY 310 GTPRLSLSEVRMDGKKEINIMVAPSGKFEELKRRAAARAAAGTPEPLAMDG 369  
DB 117 EDLKTLEFKOLKMRD-----FREILIKQ----- 140  
OY 370 GAGFTSEDRGIGITLRVAVANGIGNAKKLITKQAGE-AKYDFEIMACPGCVGGGQOP 428  
DB 141 -----NGKTVLKFAIANGFPNIGNLVQKLRKENVSNHYFEVMAACPGCINGAQOI 191  
OY 429 RSTD---KAITOKRQALVNLDEKSTLRSHENPSIRELYDTYL 469  
DB 192 RPTTGOHVRELTRKLEELYQMLP-----RSEPENSLTKHYNDPL 231  
RESULT 5  
ID AAO22158 standard; Protein; 4999 AA.  
XX  
XX AAO22158;  
XX  
XX 03-OCT-2002 (first entry)  
XX  
XX Ramoplanin biosynthetic ORF 13 protein.  
XX  
XX Ramoplanin; ramoplanin biosynthetic pathway polypeptide; antibiotic;  
KW biosynthesis gene cluster; bioengineering; peptide synthetase module;  
KW adenylation domain; hydroxyphenylglycine; HPG; antibiotic precursor;  
KW chlorinate; lipopeptide.  
XX  
XX Actinoplanes sp.

XX  
XX MO200231155-A2.  
PN  
XX  
XX 18-APR-2002.  
PD  
XX  
XX 15-OCT-2001; 2001MO-CA01462.  
PF  
XX  
XX 13-OCT-2000; 2000US-239924P.  
PR 12-APR-2001; 2001US-283296P.  
PR 24-JUL-2001; 2001US-0910813.  
XX  
XX (ECOP-) ECOPIA BIOSCIENCES INC.  
PA Farnet CM, Zazopoulos E, Staffa A;  
PI WPI; 2002-435445/46.  
DR N-PSDB; AAL40781.  
XX  
XX Novel isolated ramoplanin biosynthetic pathway polypeptide useful for  
PT chemically modifying biological molecule that is a substrate for a  
PT polypeptide encoded by a ramoplanin biosynthesis gene cluster -  
XX  
XX Claim 14; Page 153-169; 212pp; English.  
XX  
XX The invention relates to an isolated ramoplanin biosynthetic pathway  
CC polypeptide selected from a polypeptide of open reading frames (ORF)  
CC 1-32. The isolated polypeptides are useful for chemically modifying a  
CC biological molecule that is a substrate for a polypeptide encoded by a  
CC ramoplanin biosynthesis gene cluster, by contacting the biological  
CC molecule with the isolated polypeptide, where the polypeptide chemically  
CC modifies the biological molecule. The method comprises contacting the  
CC biological molecule with at least two different polypeptides encoded by  
CC ramoplanin ORFs 1-31. The polypeptides are useful for directing the  
CC biosynthesis of the antibiotic ramoplanin in microorganisms. An isolated  
CC gene cluster comprising the ORFs is useful as a substrate for  
CC bioengineering of antibiotic structures. An isolated polypeptide or its  
CC encoding nucleic acid sequence is useful for generating derivatives of  
CC ramoplanin, for improving production or for producing variants of other  
CC antibiotics of the peptide class. The isolated polypeptides are useful  
CC for synthesis of ramoplanin in vivo or in vitro, as an adenylation domain  
CC in conjunction with other peptide synthetase modules and allowing the  
CC incorporation of the into a peptide antibiotic precursor, for modifying  
CC fatty acid structure and/or enhancing fatty acid incorporation into the  
CC peptide antibiotic structure, for production of an hydroxyphenylglycine  
CC (HPG)-containing peptide antibiotic, for enhancing secretion of  
CC ramoplanin or its variants and derivatives, for enhancing uptake of  
CC precursors for ramoplanin biosynthesis, for enhancing production of  
CC ramoplanin products or its variants or derivatives, to chlorinate HPG of  
CC a peptide antibiotic precursor, and for designing specific nucleotide  
CC probes and primers for identifying and isolating putative lipopeptide  
CC -producing microorganisms. This sequence represents one of the ORF  
CC proteins of the ramoplanin producing Actinoplanes sp. microorganism of  
CC the invention.  
XX  
XX Sequence 4999 AA;  
SQ  
Query Match 5.6%; Score 144; DB 23; Length 4999;  
Best Local Similarity 21.3%; Pred. NO. 0.0017;  
Matches 133; Conservative 58; Mismatches 206; Indels 228; Gaps 27;  
OY 29 PLAASVVALATLEAPRRIGNVACAAPAAEA-----PLSHVQALAL 73  
DB 4380 PLARDLAAYAAARIDATAPALG-----ALFVQYADVADLQORVLSGSEHPDSYTSOQVA 4433  
OY 74 -----ELAKPKDPTTRKH-VQOVAVPAVVAIAETL---GLAPGATTP 112  
DB 4434 YWRQLAGVPEELDLPVDRAPAPASHRGTRVEAPVAVHHQJAEIARRNGVVFMTVO 4493  
OY 113 KQALEGRRLG-----FDEVEDFTLG-----ADLTMEGSELLARL 149  
DB 4494 TALAVLSKLAGADIPIGVAVAGRTPTDNLNLGFPVNTLVATDLGNTITDILHRT 4553  
OY 150 TE-HLEAHPHSDPELPMFTSCCPGIMALEKSYDPLIPYSSCKSPOMLAAMVKSILAE 208

```

Db      4554 RDTTHAFTHQDVP-----PEKLVEDLAPTRSLARHPLFQVMATLQASAD 4599
OY      209 KGKIAKPKDVMVSIIMPCRKQSEADRMVCVADPTLRD---DHVITVELGNITKENG 265
Db      4600 EEPFLAAGL-RVYDLPAGETPAK-----VDLDLHLHVGKDHATL----- 4641
OY      266 ILIAELPEGEMDNPMKVGSGAGVLEGTGVMERALRTAYELLETPPLRPLSLSEVRGMD 325
Db      4642 LKAADLFEGE-----TVRALADRLRT-LEMAAAPDRDLRIEVLSPG 4684
OY      326 G-----IKETNITMVPAPSGKFEELKHNRAAEAAAHGTPGLANDGAGFTSEDRGG 381
Db      4685 ERSRLLEVENDTARPVSESSVPALFAEQYAAAPDAVAVVGEQ-VSWT-----YRELDARSD 4739
OY      382 ILRLVAVANGLNNAKLITKMO-----AG-----EAKDYVEIMA 416
Db      4740 ALARSLVAAAGVSESVVVALEKSPFVLSAFLAAVAKAGVFPVDSLMPQARVDV-VAD 4798
OY      417 CPAG-----CVGGGGGPRSTDKAITOKRQ-----AALYNLDEKSTL 452
Db      4799 CGARLAVADRPMSGTLVVASAGGDSAVVSGDLTADRRAVVLPAQVPFGAAVYMTSGST 4858
OY      453 RR-----SHEN-----PSTRELYD----- 466
Db      4859 GSRKGVVTHQNLVDLATDTGWCPTPRVLFHAPHAFDASSYEIWPDLNGGVVAVPARRS 4918
OY      467 ---TYLGEPLGHKAHELLHTHYVAG 488
Db      4919 IDATVLRDLIG-AHELTTHVHTAG 4941

```

## RESULT 6

AAW89272  
ID AAW89272 standard; Protein; 748 AA.

AAW89272:

04-MAR-1999 (first entry)

Granulocytic Ehrlichia protein clone S2.

Granulocytic ehrlichiosis; Ehrlichia sp.; GE protein; infection; tick;  
diagnosis; vaccine; antigenic protein; antibody; immune response.

Ehrlichia sp.

WO9849313-A2.

05-NOV-1998.

24-APR-1998: 98WO-US08265.

25-APR-1997: 97US-0044933.

(AQUI-) AQUILA BIOPHARMACEUTICALS INC.

Beltz GA, Coughlin RT, Murphy CA, Storey J;

WPI: 1999-009432/01.  
N-PSDB: AAW82787.

New nucleic acid from the human granulocytic ehrlichiosis agent -  
PT and related antigenic proteins, vectors, transformed cells and  
antibodies, useful for diagnosis and in protective vaccines

Claim 16; Fig 7; 154pp; English.

The present sequence represents a granulocytic ehrlichia (GE) protein.  
CC nucleic acids, vectors and host cells are used for the recombinant  
CC production of GE proteins, and also in research to further characterise  
CC the proteins. GE protein-encoding nucleic acid molecules are detected  
CC by hybridisation to GE nucleic acid fragments or by using the fragments

as primers for polymerase chain reaction (PCR) amplification. GE  
CC proteins, their immunogenic fragments, and GE nucleic acid molecules  
CC encoding them are used to generate an immune response against GE.  
CC specifically as (genetic) vaccines, especially to control ehrlichiosis  
CC in humans and dogs, but also to raise Ab and to study DNA-protein  
CC interactions. Ab are used to detect GE proteins by forming an immune  
CC complex in standard assays, and correspondingly GE proteins can detect  
CC specific antibodies, especially for diagnosis, assessment and prognosis  
CC of GE infection, or of contamination of biological samples with GE.

Sequence 748 AA:

Query Match 4.9%; Score 125.5; DB 20; Length 748;

Best Local Similarity 23.2%; Pred. No. 0.0055;

Matches 66; Conservative 33; Mismatches 104; Indels 81; Gaps 12;

```

OY      248 LDHVITVELGNITKENGINLAELPGEEMDNPMKVGSGAGVLEGTGVMERALRTAYEL 307
Db      98 LEHLITV-EVIVNEE-----ITPEGK-----KTLTLEALTSGRYGVKALIKNSADV 146
OY      308 FTGTPPLRSL-----SEVRGMDGIKETNITMVPAPGS----- 340
Db      147 -NASPEPATITLGIGRCFGSKAIKHLKRYEAGAHINPTGSMSPPLAAVQANERASNL 205
OY      341 ---KFEELKHNRAAEAAAHGTP-----GPLAMDGAGFTSEDRGGGIT 383
Db      206 KEANKIYNFLNHRGADLSSTENHGTPALHLATAAGNHRTPMLLDDGAPATQDAGRFA 265
OY      384 LRYAVANGLNNAKLITKMOAGEAKYDVEIEMACPRAGCVGGGGOPRSTDKAITOKROAL 443
Db      266 LRIIAANGDKLEYRMIK-----KCPDSC-----QPLCSDMGDYALHE-AL 305
OY      444 Y--NLDEK---STLRSHENPSTRELYDYLGEPGLGHKAHELLH 482
Db      306 YSDNVTEKCFLLMKESRHLNSFFGDLINTPQANGDFTLH 349

```

## RESULT 7

AAV78923  
ID AAV78923 standard; Protein; 748 AA.

AAV78923:

23-MAY-2000 (first entry)

Granulocytic ehrlichia N11 protein sequence.

Granulocytic ehrlichia; granulocytic ehrlichiosis; vaccine; N11;  
prevent; treatment.

Ehrlichia sp.

WO200006744-A1.

10-FEB-2000.

23-OCT-1998: 98WO-US22512.

28-JUL-1998: 98US-0094381.

(AQUI-) AQUILA BIOPHARMACEUTICALS INC.

Murphy CI, Massung RF;

WPI: 2000-195304/17.  
N-PSDB: AAZ92240.

Novel granulocytic ehrlichia nucleic acid molecules, their polypeptides  
PT useful as vaccines for treating ehrlichiosis in mammals e.g. humans,  
PT pigs and dogs

Claim 15; Fig 24; 192pp; English.

XX



PS Claim 15; Fig 26; 192pp; English.

CC This sequence represents the granulocytic ehrlichia (GE) NY3 protein  
 CC sequence. The invention relates to 13 GE genes W11, W12, W13, W14, W1C,  
 CC NY1, W12, NY3, SWED, BOV, EQ, SLOV1, and SLOV2 isolated from 13 different  
 CC GE clones from a dog, a cow, a horse and ten humans. Granulocytic  
 CC ehrlichia is the causative agent of granulocytic ehrlichiosis, an acute  
 CC potentially fatal tick-borne infection. A vaccine comprising a GE nucleic  
 CC acid molecule or the polypeptide that it encodes, is used for producing  
 CC an immune response in a host to prevent granulocytic ehrlichiosis in an  
 CC animal. The protein sequences can be used to detect anti-GE antibodies in  
 CC an animal.

CC Sequence 748 AA;

Query Match 4.9%; Score 125.5; DB 21; Length 748;

Best Local Similarity 23.2%; Pred. No. 0.0055;

Matches 66; Conservative 33; Mismatches 104; Indels 81; Gaps 12;

OY 248 LDHVTITVELGNIFERGINLAELPEGEMDNPMVGSGVLEGTGGVMEALTRAVEL 307  
 DB 98 LEHLITT-EVIVSNEE-----ITPECK-----KTLTRREALTSGKYGVAALINKSADV 146  
 OY 308 FTGPELPLRLSL-----SEVRGMDGKETNTITWVPAPGS----- 340  
 DB 147 -NASPEPAITLIGRCRCGSKAIKHLKRVYEAHAHINPTGSMPLAAVQANAEASNL 205  
 OY 341 ----KFEELKRRARAARAAAHGTP-----GPLAMDGAGFTSDEGRGIGIT 383  
 DB 206 KEANKRIVNELLHRGADLSSTHTGTPLALHTAAGNHRTAMLLDKGAPATORDARGRTA 265  
 OY 384 LRVAANGIGNAKKILITKMGAEAKYDFEIMACPGACGCGGGRSDTKAITOKROAL 443  
 DB 266 LHIAAANGDGKIRYRIAK-----KCPDSC-----QPLCSMDGTALHE-AL 305  
 OY 444 Y--NLDER---STLRSHENPSIRELYDTYLGEPLGKRAHELLH 482  
 DB 306 YSDNTEKCFLEKMLKESRKHLNSNSFFDGLLMTPEANGDTLLH 349

RESULT 10

AAW82490  
 ID AAW82490 standard; Protein; 573 AA.

AAW82490;

AC AAW82490;

DT 02-MAR-1999 (first entry)

DE Ehrlichia sp. E82.1 protein.

XX

KM Granulocytic ehrlichia; GE; E82; tick-borne infection; fatal; vaccine;

KW Immune response; detection; diagnosis; Ehrlichiosis.

XX Ehrlichia sp.

OS W09849312-A2.

XX W09849312-A2.

PD 05-NOV-1998.

XX 24-APR-1998; 98WO-US08264.

XX 25-APR-1997; 97US-0044869.

XX (AQUIT-) AQUITA BIOPHARMACEUTICALS INC.

XX Beltez G, Coughlin RT, Murphy C, Storey J;

XX WPI; 1999-034663/03.

XX DR N-PSDB; AAV65142.

PT New isolated granulocytic ehrlichia nucleic acids - used to develop  
 PT products for use in vaccines for inhibiting Ehrlichiosis and for use  
 PT in detection and diagnosis

XX  
 PS Disclosure; Fig 10; 184pp; English.

CC This sequence encodes the E82.1 protein which is associated with and has  
 CC been isolated from HL60 cells infected with Ehrlichia sp. GE is an acute  
 CC potentially fatal tick borne infection and the proteins described in  
 CC this invention can be used in vaccines to elicit a beneficial immune  
 CC response in an animal to GE. They can be used for inhibiting Ehrlichiosis  
 CC in an animal. The products can also be used for detection and diagnosis.

CC Sequence 573 AA;

Query Match 4.8%; Score 122.5; DB 20; Length 573;

Best Local Similarity 26.1%; Pred. No. 0.0072;

Matches 57; Conservative 14; Mismatches 80; Indels 67; Gaps 9;

OY 286 AGVLFPTGGVNE---ALRTAYELFTGTPRLSLSEVRGMDGKETNTITWVPAPGSKF 342  
 DB 3 AGAHINTPTGSMSPPLAAVQANE-----ASNKEAN-----KI 36  
 OY 343 EELKRRARAARAAAHGTP-----GPLAMDGAGFTSDEGRGIGITLRVAVA 389  
 DB 37 VNFLHRGADLSSTHTGTPLALHTAAGNHRTAMLLDKGAPATORDARGRTALHIAA 96  
 OY 390 NGLGNNAKKILITKMGAEAKYDFEIMACPGACGCGGGRSDTKAITOKROALY--NLD 447  
 DB 97 NGDGKLYRIAK-----KCPDSC-----QPLCSMDGTALHE-ALYSDNVT 136  
 OY 448 EK---STLRSHENPSIRELYDTYLGEPLGKRAHELLH 482  
 DB 137 EKCFLEKMLKESRKHLNSNSFFDGLLMTPEANGDTLLH 174

RESULT 11

AAU02197  
 ID AAU02197 standard; Protein; 965 AA.

AAU02197;

DT 26-SEP-2001 (first entry)

DE Human glutamate receptor-like protein, MEM2.

XX Glutamate receptor; MEM1; therapeutic; diagnostic; MEM2;

KM human; Alzheimer's disease; Parkinson's disease; cancer; nephrology;

KW female reproductive health; lung disorder; brain disorder; schizophrenia;

KW heart disorder; arrhythmia; muscular disorder; clotting deficiency; MEM3;

KW cobalamin deficiency; pernicious anaemia; diabetes; MEM4; MEM5; MEM6;

XX vision-related disorder; neoplastic pathology; MEM7; MEM8.

OS Homo sapiens.

XX

XX Key Location/Qualifiers

FT Misc-difference 1..965

FT /note= "All X's have been input by indexer; these

XX represent unreadable residues in the specification"

XX W0200144473-A2.

XX 21-JUN-2001.

XX 14-DEC-2000; 2000WO-US33909.

XX 14-DEC-1999; 99US-0170564.

XX 27-DEC-1999; 99US-0173165.

XX 27-DEC-1999; 99US-0173162.

XX 29-DEC-1999; 99US-0173544.

XX 04-JAN-2000; 2000US-9965654.

XX 09-AUG-2000; 2000US-0223929.

XX 13-DEC-2000; 2000US-9965655.

XX (CURA-) CURAGEN CORP.

PI Spaderna SK, Quinn KE, Shinkets RA, Muralidhara P, Spytek KA;  
 XX RPI: 2001-398154/42.  
 DR N-PSDB; AAS06333.  
 XX  
 PT Novel polypeptide comprising members of protein families (e.g.,  
 PT seven-pass transmembrane receptor proteins) according to presence of  
 PT domains and sequence relatedness are useful for treating or preventing,  
 PT e.g., Alzheimer's and Parkinson's -  
 XX  
 PS Claim 1; Fig 8; 162pp; English.  
 XX  
 CC The sequence represents the amino acid sequence of glutamate receptor-  
 CC like protein, MEM2, selected from a group (MEM1-MEM8) comprising  
 CC members of protein families according to the presence of domains and  
 CC sequence relatedness, e.g., seven-pass transmembrane receptor protein  
 CC (MEM1), glutamate receptor (MEM2-MEM4), potassium channel protein (MEM5),  
 CC phosphatase I protein (MEM6), and retinol-binding protein (MEM7-MEM8).  
 CC The MEM polypeptides (I), nucleic acids (II), and antibodies (III) are  
 CC all useful for treating or preventing a pathology associated with (I)  
 CC comprising administering (I), (II), or (III) to a subject (preferably a  
 CC human). In addition, (I), (II), and (III) may be used to manufacture a  
 CC medicament for treating a syndrome associated with a human disease that  
 CC is associated with (I). Furthermore, (I) may be used to identify agents  
 CC that bind to it, screen modulators of its activity and determine the  
 CC presence or predisposition to a disease associated with altered levels of  
 CC (I). Disorders for MEM1 include Alzheimer's or Parkinson's Disease,  
 CC cancer, nephrology, and female reproductive health. Disorders for MEM4  
 CC include those involving the lung and/or brain (e.g., schizophrenia). For  
 CC MEM5, disorders include heart (arrhythmic disorders) and other muscular  
 CC disorders, clotting deficiencies and cobalamin deficiencies (e.g.,  
 CC pernicious anaemia). Such disorders for MEM6 include diabetes, whereas  
 CC disorders for MEM7 and MEM8 include vision-related disorders, cancer,  
 CC and other neoplastic pathologies.  
 CC  
 XX  
 XX Sequence 965 AA:  
 SO  
 Query Match 4.7%; Score 122; DB 22; Length 965;  
 Best Local Similarity 22.3%; Pred. No. 0.019;  
 Matches 122; Conservative 45; Mismatches 187; Indels 194; Gaps 23;

QY 439 RQALVNLDEKSTLR-----SHENSTIRELY-----DTYLG 470  
 XX ||| | : | ||| : |  
 Db 387 LFAALANGSAPRLRKCCYCTIDLERNAEDTPPEFLYLVGDKYTGALRGRWGLVG 446  
 QY 471 EPLGKRAH 478  
 : | : ||  
 Db 447 DLLAGRAH 454  
 RESULT 12  
 AA02198  
 ID AA02198 standard; Protein; 971 AA.  
 XX  
 AC AA02198;  
 XX  
 DT 26-SEP-2001 (first entry)  
 XX  
 XX Human glutamate receptor-like protein, MEM3.  
 XX  
 KM Glutamate receptor; MEM1; therapeutic; diagnostic; MEM2;  
 KM human; Alzheimer's disease; Parkinson's disease; cancer; nephrology;  
 KM female reproductive health; lung disorder; brain disorder; schizophrenia;  
 KM heart disorder; arrhythmia; muscular disorder; clotting deficiency; MEM3;  
 KM cobalamin deficiency; pernicious anaemia; diabetes; MEM4; MEM5; MEM6;  
 KM vision-related disorder; neoplastic pathology; MEM7; MEM8.  
 OS Homo sapiens.  
 XX  
 PN WO200144473-A2.  
 XX  
 PD 21-JUN-2001.  
 XX  
 PD 14-DEC-2000; 2000MO-US33909.  
 PF  
 XX 14-DEC-1999; 99US-0170564.  
 PR 27-DEC-1999; 99US-0173165.  
 PR 27-DEC-1999; 99US-0173362.  
 PR 29-DEC-1999; 99US-0173544.  
 PR 04-JAN-2000; 2000US-9966564.  
 PR 09-AUG-2000; 2000US-0223929.  
 PR 13-DEC-2000; 2000US-9966565.  
 PR  
 XX (CURA-) CURAGEN CORP.  
 XX  
 PI Spaderna SK, Quinn KE, Shinkets RA, Muralidhara P, Spytek KA;  
 XX RPI: 2001-398154/42.  
 DR N-PSDB; AAS06334.  
 XX  
 PT Novel polypeptide comprising members of protein families (e.g.,  
 PT seven-pass transmembrane receptor proteins) according to presence of  
 PT domains and sequence relatedness are useful for treating or preventing,  
 PT e.g., Alzheimer's and Parkinson's -  
 XX  
 PS Claim 1; Fig 13; 162pp; English.  
 XX  
 CC The sequence represents the amino acid sequence of glutamate receptor-  
 CC like protein, MEM3, selected from a group (MEM1-MEM8) comprising  
 CC members of protein families according to the presence of domains and  
 CC sequence relatedness, e.g., seven-pass transmembrane receptor protein  
 CC (MEM1), glutamate receptor (MEM2-MEM4), potassium channel protein (MEM5),  
 CC phosphatase I protein (MEM6), and retinol-binding protein (MEM7-MEM8).  
 CC The MEM polypeptides (I), nucleic acids (II), and antibodies (III) are  
 CC all useful for treating or preventing a pathology associated with (I)  
 CC comprising administering (I), (II), or (III) to a subject (preferably a  
 CC human). In addition, (I), (II), and (III) may be used to manufacture a  
 CC medicament for treating a syndrome associated with a human disease that  
 CC is associated with (I). Furthermore, (I) may be used to identify agents  
 CC that bind to it, screen modulators of its activity and determine the  
 CC presence or predisposition to a disease associated with altered levels of  
 CC (I). Disorders for MEM1 include Alzheimer's or Parkinson's Disease,  
 CC cancer, nephrology, and female reproductive health. Disorders for MEM4  
 CC include those involving the lung and/or brain (e.g., schizophrenia). For

CC MEM5, disorders include heart (arrhythmic disorders) and other muscular  
 CC disorders, clotting deficiencies and cobalamin deficiencies (e.g.,  
 CC pernicious anemia). Such disorders for MEM6 include diabetes, whereas  
 CC disorders for MEM7 and MEM8 include vision-related disorders, cancer,  
 CC and other neoplastic pathologies.

XX sequence 971 AA;

Query Match 4.7%; Score 122; DB 22; Length 971;

Best Local Similarity 21.9%; Pred. No. 0.019;

Matches 120; Conservative 46; Mismatches 189; Indels 192; Gaps 22;

7 KPCAAVSISSGSCRAQVAPRAPLASTVVALATLEAPARRLGNVA--CAAAAPAAEAP 64  
 DB 25 QPCGVLARIGSVRLGALLPRAPLARAARALAPRLPHNISLELVAAAPRDP 84  
 65 LSHVQ---QAL-----AEIAKPKDDP--TRKHVCQVAPAVVAIAETGLAP-GATTP 112  
 DB 85 ASLTRGLQALVPPGVAALLAPPEARPELLQLHFLAAATETPVLSTLRREARAPLAGAPNP 144  
 113 KQALEGRRLGDEVDITFGADLTMEGSELLHNLTEHLNHPHSDEPLPFTSCCPG 172  
 DB 145 FHL-----QLHMASPLET-----LLDVLAVVLQAHAMEDVGLACRTQDDG 185  
 173 WIAMLEKSY----PDLTIPYVSSCKSPOMLAAMVKSYLEKKGIAPKDMVMVSIIMPCTR 228  
 DB 186 GLVALMTSRAGRPOLVLDLSRDTGDAGRLAPMAAPVSGEAP----- 231  
 229 QSEADRMFCVDADPTLRDLNHYITTVELG-NIFKRGINLAELPGEENDMPGVSGAG 287  
 DB 232 -----VPAVLVLCGLDIAARRARYLEAVPPGPHW----- 258  
 288 VLEGTGGVMEALRTAYELFTGTPLRLSLSEVR-----GMDGK 328  
 DB 259 -LLGTP--LPPKALPTA-----GLPPGLALGCVARPPLEAIIHYQIVARALGSAQV 310  
 329 ETNITWVAP-----GSKFEELKRAAARAALAHGTGPLAMDGAGFTSEGRGG 381  
 DB 311 QPKRALLPAPVNGDLOPAGPESPPGRFLRLANTSFQGRTPV-WVTGSS-PDEGQ-- 366  
 382 TILRAVAVANGLNAKLITRKMAQGEAKYDFVEIMACPG--CVGGGGQPPSTDKAITQKR 439  
 DB 367 -----CPAGOLCL-----DPGTNDSATLDA 387  
 440 QALVNLDEKSTLRR-----SHENPSIRELY-----DYLGE 471  
 DB 388 FALLANGSAPRALRKCCGYCIDLERLEADTPDFEELYLVGDGKYGALRDGRWTGLVGD 447  
 472 PLGHRKAH 478  
 DB 448 LLAGRAH 454

RESULT 13  
 AAO18086  
 ID AAO18086 standard; Protein; 901 AA.  
 AC AAO18086;  
 XX  
 XX 05-SEP-2002 (first entry)  
 XX  
 XX Human N-methyl-D-aspartate receptor.  
 DE Human N-methyl-D-aspartate receptor.  
 XX  
 XX Human; N-methyl-D-aspartate receptor; NMDA receptor; asthma;  
 KW Chromosome 19p13.3; genito-urinary system disorder; urinary incontinence;  
 KW benign prostate hyperplasia; nervous system disorder; anastomatic;  
 KW utrophic; noctropic; neuroprotective; gene therapy.  
 XX  
 OS Homo sapiens.  
 XX  
 XX WO200240538-A2.  
 XX  
 XX 23-MAY-2002.  
 PD

XX 16-NOV-2001; 2001WO-EP13264.  
 PF  
 XX 17-NOV-2000; 2000US-249273P.  
 PR  
 XX (FARB ) BAYER AG.  
 PA  
 XX  
 XX  
 PI  
 XX  
 XX  
 DR WPI; 2002-500204/53.  
 DR N-PSDB; AAIL7419.  
 XX

XX New human N-methyl-D-aspartate receptor polypeptide for identifying  
 XX modulating agents useful in treating diseases e.g. asthma

PS Claim 25; Fig 2; 108pp; English.  
 CC The present invention provides the protein and coding sequences of the  
 CC human N-methyl-D-aspartate (NMDA) receptor. The sequences are useful in  
 CC the identification of modulators of NMDA receptor, and in the treatment  
 CC of diseases such as asthma, a genito-urinary system disorder such as  
 CC urinary incontinence and benign prostate hyperplasia or a peripheral or  
 CC central nervous system disorder. The NMDA gene is found on chromosome  
 CC 19p13.3. The present sequence is the protein of the invention.

XX SO Sequence 901 AA;

Query Match 4.5%; Score 116; DB 23; Length 901;

Best Local Similarity 21.7%; Pred. No. 0.067;

Matches 128; Conservative 49; Mismatches 207; Indels 206; Gaps 25;

7 KPCAAVSISSGSCRAQVAPRAPLASTVVALATLEAPARRLGNVA--CAAAAPAAEAP 64  
 DB 25 QPCGVLARIGSVRLGALLPRAPLARAARALAPRLPHNISLELVAAAPRDP 84  
 65 LSHVQ---QAL-----AEIAKPKDDP--TRKHVCQVAPAVVAIAETGLAP-GATTP 112  
 DB 85 ASLTRGLQALVPPGVAALLAPPEARPELLQLHFLAAATETPVLSTLRREARAPLAGAPNP 144  
 113 KQALEGRRLGDEVDITFGADLTMEGSELLHNLTEHLNHPHSDEPLPFTSCCPG 172  
 DB 145 FHL-----QLHMASPLET-----LLDVLAVVLQAHAMEDVGLACRTQDDG 185  
 173 WIAMLEKSY----PDLTIPYVSSCKSPOMLAAMVKSYLEKKGIAPKDMVMVSIIMPCTR 228  
 DB 186 GLVALMTSRAGRPOLVLDLSRDTGDAGRLAPMAAPVSGEAP----- 231  
 229 QSEADRMFCVDADPTLRDLNHYITTVELG-NIFKRGINLAELPGEENDMPGVSGAG 287  
 DB 232 -----VPAVLVLCGLDIAARRARYLEAVPPGPHW----- 258  
 288 VLEGTGGVMEALRTAYELFTGTPLRLSLSEVR-----GMDGK 328  
 DB 259 -LLGTP--LPPKALPTA-----GLPPGLALGCVARPPLEAIIHYQIVARALGSAQV 310  
 329 ETNITWVAP-----GSKFEELKRAAARAALAHGTGPLAMDGAGFTSEGRGG 381  
 DB 311 QPKRALLPAPVNGDLOPAGPESPPGRFLRLANTSFQGRTPV-WVTGSS-PDEGQ-- 366  
 382 TILRAVAVANGLNAKLITRKMAQGEAKYDFVEIMACPG--CVGGGGQPPSTDKAITQKR 439  
 DB 367 -----CPAGOLCL-----DPGTNDSATLDA 387  
 440 QALVNLDEKSTLRR-----SHENPSIRELY-----DYLGE 471  
 DB 388 FALLANGSAPRALRKCCGYCIDLERLEADTPDFEELYLVGDGKYGALRDGRWTGLVGD 447  
 472 PLGHRKAH 478  
 DB 448 LLAGRAH 454

RESULT 14  
 477 YGICIDLERLEADTPDFEELYLVGDGKYGALRDGRWTGLVGDLLAGRAH 526

AA02199 standard; Protein: 1043 AA.  
 ID AA02199 standard; Protein: 1043 AA.  
 AC AA02199;  
 XX  
 XX  
 DT 26-SEP-2001 (first entry)  
 XX  
 XX  
 DE Human glutamate receptor-like protein, MEM4.  
 XX  
 XX Glutamate receptor; MEM1: therapeutic; diagnostic; MEM2:  
 KM human; Alzheimer's disease; Parkinson's disease; cancer; nephrology;  
 KM female reproductive health; lung disorder; brain disorder; schizophrenia;  
 KM heart disorder; arrhythmia; muscular disorder; clotting deficiency; MEM3;  
 KM cobalamin deficiency; pernicious anaemia; diabetes; MEM4; MEM5; MEM6;  
 KM vision-related disorder; neoplastic pathology; MEM7; MEM8.  
 XX  
 OS Homo sapiens.  
 XX  
 PN W0200144473-A2.  
 PD 21-JUN-2001.  
 XX  
 PF 14-DEC-2000; 2000MO-US33909.  
 XX  
 XX 14-DEC-1999; 99US-0170564.  
 PR 27-DEC-1999; 99US-0173165.  
 PR 27-DEC-1999; 99US-0173362.  
 PR 29-DEC-1999; 99US-0173544.  
 PR 04-JAN-2000; 2000US-9966564.  
 PR 09-AUG-2000; 2000US-0223929.  
 PR 13-DEC-2000; 2000US-9966565.  
 XX  
 XX (CURA-) CURAGEN CORP.  
 PA  
 PI Spaderna SK, Quinn KE, Shinkets RA, Muralidhara P, Spytek KA;  
 XX  
 DR WPI; 2001-398154/42.  
 DR N-PSDB; AAS06335.  
 XX  
 XX Novel polypeptide comprising members of protein families (e.g.,  
 PT seven-pass transmembrane receptor proteins) according to presence of  
 PT domains and sequence relatedness are useful for treating or preventing,  
 PT e.g., Alzheimer's and Parkinson's -  
 XX  
 PS Claim 1; Fig 18; 162pp; English.  
 XX  
 XX The sequence represents the amino acid sequence of glutamate receptor-  
 CC like protein, MEM4, selected from a group (MEM1-MEM6) comprising  
 CC members of protein families according to the presence of domains and  
 CC sequence relatedness, e.g., seven-pass transmembrane receptor protein  
 CC (MEM1), glutamate receptor (MEM2-MEM4), potassium channel protein (MEM5),  
 CC phosphatase I protein (MEM6), and retinol-binding protein (MEM7-MEM8).  
 CC The MEM polypeptides (I), nucleic acids (II), and antibodies (III) are  
 CC all useful for treating or preventing a pathology associated with (I)  
 CC comprising administering (I), (II), or (III) to a subject (preferably a  
 CC human). In addition, (I), (II), and (III) may be used to manufacture a  
 CC medicament for treating a syndrome associated with a human disease that  
 CC is associated with (I). Furthermore, (I) may be used to identify agents  
 CC that bind to it, screen modulators of its activity and determine the  
 CC presence or predisposition to a disease associated with altered levels of  
 CC (I). Disorders for MEM1 include Alzheimer's or Parkinson's Disease,  
 CC cancer, nephrology, and female reproductive health. Disorders for MEM4  
 CC include those involving the lung and/or brain (e.g., schizophrenia). For  
 CC MEM5, disorders include heart (arrhythmic disorders) and other muscular  
 CC disorders, clotting deficiencies and cobalamin deficiencies (e.g.,  
 CC pernicious anaemia). Such disorders for MEM6 include diabetes, whereas  
 CC disorders for MEM7 and MEM8 include vision-related disorders, cancer,  
 CC and other neoplastic pathologies.  
 XX  
 SO Sequence 1043 AA;

Query Match 4.5%; Score 116; DB 22; Length 1043;  
 Best Local Similarity 21.7%; Pred. No. 0.085;

Matches 128; Conservative 49; Mismatches 207; Indels 206; Gaps 25;  
 QY 7 KPCAAYSIRSCSAROVAPRAPLAATVVALATEAPARRGNYA--CAAAAPAAEAP 64  
 DB 25 QPCGVALARLGGSVRLGALPRAPLARARAAALARALRRLPHNLSLEYVAAPRRDP 84  
 QY 65 LSHVQ---QAL-----AELAKPRDDP--TRKHVCQVAVAVVAIAETLGLAP-GATTP 112  
 DB 85 ASLTRGICQALVPCVAVALLAFPEARPELLQLHFLAAETPEVLSLLRREARAPLAPNP 144  
 QY 113 KQLABEIRLGFPEVDITLFGADLTMEBSSELHNLTELEAHNPSDELPMTSCCPG 172  
 DB 145 FHL-----QLHMASPLET-----LLDYLVAVLQAHMEDVGLALCRQDPG 185  
 QY 173 WIMLEKSY-----PDILPYVSSCKSPOMLAAMVKSYLEAKKGIAPKDMVMSIMPTCK 228  
 DB 186 GLVALMTSRAGRPQVLYLDSRDTGDAGLRALAPMAAPVGEAP----- 231  
 QY 229 QSEADRDWFCVDAPTLRQLDHYITVVELG--NIFKRGIMIAELPEGEMDNPKVSGSAG 287  
 DB 232 -----VPAAVLIGCDIARARRVLEAVPPGPHW----- 258  
 QY 288 VLFETTGVMNALRTAYELFTGTPLPRLSLEYR-----GMDGTR 328  
 DB 259 -LIGTP--LPPKALPTA---GLPPGLALGEVAPRPLEMAIHDIYQLVARALGSAQV 310  
 QY 329 ETNITWPPAP-----GSKFEELKHRAA 351  
 DB 311 QPKRALLPAPVNGCDLPACPESPGRFLANFLANTSQGTGVTGWTTGSSQVIMSHFKV 370  
 QY 352 ARAEAAAHCTPGPLADMGAGFTSEDR-----GGITLRVAVANGLGANAKL--ITKQA 404  
 DB 371 WSLRDRPRGAP---AAATVGSW--RDGQDLERGGASARPPPGQAQWPKLVTITLE- 424  
 QY 405 GEAKYDFV-----ETMACPAG--CVGGGQCPSTDKATQKQAAALYNIDESTLRR-- 454  
 DB 425 ---HPFVFARDDPEGQCPAGQCL---DPGTNDSATLDALFAALANSAPRALRKCC 476  
 QY 455 -----SHENPSIRELY-----DYLGEPHGKHAH 478  
 DB 477 YGICIDLEHLAEIDTPDFELIYVGDKYCALDKGRNTGLVGDLAGRAH 526

RESULT 15  
 AA02159 standard; Protein: 4999 AA.  
 ID AA02159 standard; Protein: 4999 AA.  
 AC AA02159;  
 XX  
 XX  
 DT 03-OCT-2002 (first entry)  
 XX  
 XX  
 DE Ramoplanin biosynthetic ORF 14 protein.  
 XX  
 XX Ramoplanin; ramoplanin biosynthetic pathway polypeptide; antibiotic;  
 KM Ramoplanin biosynthesis gene cluster; bioengineering; peptide synthetase module;  
 KM adenylation domain; hydroxyphenylglycine; HPG; antibiotic precursor;  
 KM chlorinate; lipopeptide.  
 XX  
 OS Actinoplanes sp.  
 XX  
 PN W0200231155-A2.  
 PD 18-APR-2002.  
 XX  
 PF 15-OCT-2001; 2001MO-CA01462.  
 XX  
 PR 13-OCT-2000; 2000US-239924P.  
 PR 12-APR-2001; 2001US-283296P.  
 PR 24-JUL-2001; 2001US-0910813.  
 XX  
 PA (ECOP-) ECOPRIA BIOSCIENCES INC.  
 XX  
 PI Farnet CM, Zazopoulos E, Stalfa A;

Search completed: June 3, 2003, 16:07:35  
JOB time : 75 secs

```
XX WPI: 2002-435445/46.
DR N-PSDB: AAL40781.
XX
PT Novel isolated ramoplanin biosynthetic pathway polypeptide useful for
PT chemically modifying biological molecule that is a substrate for a
PT polypeptide encoded by a ramoplanin biosynthesis gene cluster -
XX
PS Claim 14: Page 169-186; 212pp; English.
XX
CC The invention relates to an isolated ramoplanin biosynthetic pathway
CC polypeptide selected from a polypeptide of open reading frames (ORF)
CC 1-33. The isolated polypeptides are useful for chemically modifying a
CC biological molecule that is a substrate for a polypeptide encoded by a
CC ramoplanin biosynthesis gene cluster, by contacting the biological
CC molecule with the isolated polypeptide, where the polypeptide chemically
CC modifies the biological molecule. The method comprises contacting the
CC biological molecule with at least two different polypeptides encoded by
CC ramoplanin ORFs 1-31. The polypeptides are useful for directing the
CC biosynthesis of the antibiotic ramoplanin in microorganisms. An isolated
CC gene cluster comprising the ORFs is useful as a substrate for
CC bioengineering of antibiotic structures. An isolated polypeptide or its
CC encoding nucleic acid sequence is useful for generating derivatives of
CC ramoplanin, for improving production or for producing variants of other
CC antibiotics of the peptide class. The isolated polypeptides are useful
CC for synthesis of ramoplanin in vivo or in vitro, as an adenylation domain
CC in conjunction with other peptide synthetase modules and allowing the
CC incorporation of Thr into a peptide antibiotic precursor, for modifying
CC fatty acid structure and/or enhancing fatty acid incorporation into the
CC peptide antibiotic structure, for production of an hydroxyphenylglycine
CC (HPG)-containing peptide antibiotic, for enhancing secretion of
CC ramoplanin or its variants and derivatives, for enhancing uptake of
CC precursors for ramoplanin biosynthesis, for enhancing production of
CC ramoplanin products or its variants or derivatives, to chlorinate HPG of
CC a peptide antibiotic precursor, and for designing specific nucleotide
CC probes and primers for identifying and isolating putative lipopeptide
CC -producing microorganisms. This sequence represents one of the ORF
CC proteins of the ramoplanin producing Actinoplanes sp. microorganism of
CC the invention.
XX
SQ Sequence 4999 AA:
Query Match 4.5%; Score 116; DB 23; Length 4999;
Best Local Similarity 24.1%; Pred. No. 1.1;
Matches 100; Conservative 32; Mismatches 171; Indels 112; Gaps 20;
OY 8 PCAAVSIRGSSCRAROVA---PRAPLASTVRAVALTEAPARRLGNVACAAAPRAEAP 64
DB 4400 PSDAISQVVAHNR-RQLAGAPDELPLPADHPPRAETVR-----GHTVEFTVPPA---- 4448
OY 65 LSHVQOALAEIAKPKDDPTKRHVQVAAVRAVAIETG-LAPGATTPKQLAEGLRIG 123
DB 4449 ---VHHQIAELARRNG-----VTVEFTVQALAVLLSKLAGTDIRIGAVAGRT-- 4495
OY 124 FDEVEFDTLFG-----ADLTMEGSELHRLTE-HLEAHPHSEDEPLMFTSCPGW 173
DB 4496 -DPTIDNLGFFVNTLVLTDLTGAPNTITDLHRTDTLHAFTHODV----- 4543
OY 174 IAMEKSYPDLLIPYSSKSP--QNMILAAVKSYLEKKGIAPKDMVNVSIMPTCKOSE 231
DB 4544 ---FEKLVEDLAPTRSLARHPLFQVWMTLQSTGRAGEAAELPGLERAVLSPGGVAAK--- 4597
OY 232 ARDMFCVADPTLRQLDHNVTITVELGNIFKERGINLAELPEGEMDPMGCVSGAGVLF 291
DB 4598 -----VDLD-----LSISEAYDDDG-----RPAGL---AGTLVA 4623
OY 292 TTG-----GYMEAALRTAYELFTGTP-LRLSLSEVGMKG-----IKETNITWPA 340
DB 4624 AADLPEHGTAERAGTARLAVLVPADPARLGDVDLDGEERRLVLTGNDTTAAVPAV 4683
OY 341 KFEELKRRRAARAERAAGHTGPIPLMDGA--GFTSEDRGSGITLRVAVANGIG 393
DB 4684 AVPELIERRAARAEPEEGA-----VMCGDTHLRVGEELNARNRRLRLIVERGAG 4731
```

**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 3, 2003, 16:09:12 ; Search time 42 Seconds

(without alignments)  
1137.591 Million cell updates/sec

Title: CAC80065

Perfect score: 2576  
Sequence: 1 MSALVKPCAASVIRSSCR.....HELTHTHYVAGVEEKDEKK 497

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database :  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1010.5	39.2	608	2 C72405	hydrogenase (EC 1.1.1.1)
2	877	34.0	585	2 D57150	hydrogenase (EC 1.1.1.1)
3	824	32.0	1206	2 T18557	probable hydrogenase (EC 1.1.1.1)
4	794.5	30.8	606	2 S13526	hydrogenase (EC 1.1.1.1)
5	743.5	28.9	574	1 HQCLIP	hydrogenase (EC 1.1.1.1)
6	741	28.8	582	2 JG6002	hydrogenase (EC 1.1.1.1)
7	741	28.8	582	2 D96803	hydrogenase (EC 1.1.1.1)
8	633	24.6	421	1 HQDYFL	cytochrome-c3 hydrolase (EC 1.1.1.1)
9	599.5	23.3	421	1 HQDYLV	cytochrome-c3 hydrolase (EC 1.1.1.1)
10	512.5	19.9	645	2 G72256	hydrogenase (EC 1.1.1.1)
11	311.5	12.1	538	2 T40992	hydrogenase (EC 1.1.1.1)
12	255	9.9	491	2 S63206	hydrogenase (EC 1.1.1.1)
13	245.5	9.5	450	2 B97297	hydrogenase (EC 1.1.1.1)
14	232	9.0	469	2 S31336	hydrogenase (EC 1.1.1.1)
15	138.5	5.4	123	1 HQDYFS	hydrogenase (EC 1.1.1.1)
16	136.5	5.3	301	2 B72256	hydrogenase (EC 1.1.1.1)
17	130	5.0	124	1 HQDYVS	hydrogenase (EC 1.1.1.1)
18	125.5	4.9	748	2 T08612	hydrogenase (EC 1.1.1.1)
19	117	4.5	24	2 S33302	hydrogenase (EC 1.1.1.1)
20	114.5	4.4	884	2 P06067	hydrogenase (EC 1.1.1.1)
21	114.5	4.4	2588	2 T14342	hydrogenase (EC 1.1.1.1)
22	112.5	4.4	913	2 B73362	hydrogenase (EC 1.1.1.1)
23	112.5	4.4	913	2 AB2587	hydrogenase (EC 1.1.1.1)
24	112.5	4.3	755	2 S23441	hydrogenase (EC 1.1.1.1)
25	111.5	4.3	891	1 JN0867	hydrogenase (EC 1.1.1.1)
26	110.5	4.3	451	2 C81059	hydrogenase (EC 1.1.1.1)
27	110	4.3	503	2 F95958	hydrogenase (EC 1.1.1.1)
28	109.5	4.3	857	1 D35905	hydrogenase (EC 1.1.1.1)
29	109.5	4.3	857	2 G91060	hydrogenase (EC 1.1.1.1)

30	109.5	4.3	861	2 E85905	heat shock protein
31	109.5	4.3	4735	2 T17463	ribosecyclic polyketide synthase
32	109	4.2	6420	2 T30283	polyketide synthase
33	108.5	4.2	594	1 G81847	dihydroxyacetone diphosphate
34	108.5	4.2	1867	2 S61703	fatty acid synthase
35	108	4.2	5149	2 F83345	probable non-ribosomal
36	107.5	4.2	391	2 T35574	probable alanine r
37	107	4.2	378	2 B75547	iron-sulfur cofact
38	107	4.2	458	2 G64507	hypothetical prote
39	106	4.1	1772	2 T36105	probable large gly
40	106	4.1	1885	1 J04086	fatty acid synthas
41	105.5	4.1	3491	2 T43231	probable 6-deoxyer
42	105	4.1	896	2 G81709	translation initia
43	105	4.1	1613	2 T06678	hypothetical prote
44	105	4.1	1621	2 A82255	hypothetical prote
45	104.5	4.1	733	2 T35429	probable fatty ox

## ALIGNMENTS

RESULT 1	
C72405	hydrogenase (EC 1.18.99.1) (Fe) large chain [similarity] - Thermotoga maritima (strain C:Species: Thermotoga maritima
C:Date: 11-Jun-1999	#sequence_revision 11-Jun-1999 #ext_change 21-Jul-2000
C:Accession: C72405	R:Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; Bric Garret, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, C.M.
Nature 399, 323-329, 1999	
A:Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome	
A:Reference number: A72200; PMID:99287316; PMID:10360571	
A:Accession: C72405	
A:Status: preliminary	
A:Molecule type: DNA	
A:Residues: 1-608 <ARN>	
A:Cross-references: GB:AE001705; GB:AE000512; NID:94980694; PIDN:AMD35293.1; PID:9498	
A:Experimental source: strain MS88	
C:Genetics:	
A:Gene: TM0201	
C:Superfamily: hydrogenase (Fe) large chain; ferredoxin 2(4Fe-4S) homology	
C:Keywords: 2Fe-2S; 4Fe-4S; iron; metalloprotein; oxidoreductase	
F:62,74,77,91/Binding site: 2Fe-2S cluster (Cys) (covalent) #status predicted	
F:123,127,130,136/Binding site: 4Fe-4S cluster (His, Cys, Cys, Cys) (covalent) (type	
F:174,177,180,227/Binding site: 4Fe-4S cluster (Cys) (covalent) #status predicted	
F:184,217,220,223/Binding site: 4Fe-4S cluster (Cys) (covalent) #status predicted	
F:329,384,528,532/Binding site: 4Fe-4S cluster (Cys) (covalent) #status predicted	
F:532/Binding site: diron cofactor (Cys) #status predicted	
Query Match	39.2% Score 1010.5; DB 2; Length 608;
Best Local Similarity	47.9% Pred. No. 1.9e-63;
Matches 216; Conservative	56; Mismatches 112; Indels 67; Gaps 8;
OY	49 LGNVCAAAP-AAEAPLSIVQOALAEAPKPDPTKRNHGVAVPAVAITGLAP 107
DB	218 ICGGCAAFPTGAIYENSAKVYLEEKE-----KLVVPTASVRAIIEERYAP 272
OY	108 GATTPQALAEGLRLGFDEVDLTFGADLTMEGSELRLRTEHLEAPHSQ-EPLPMF 166
DB	273 GTISTGQVAAALRRIGDYDFDTNFGADLTMEGSEFRLERK-----GDELEPMF 325
OY	167 TSCCPGRIAMLESTYDLITVYSSCKSPOMLAAMKSYLAEEKGIAPKPMVSTPCT 226
DB	326 TSCCPGWNILVERVPELRLRSSASPOQMLSAMKTYFAELGVAPEDIFVSTIMPCT 385
OY	227 RKQSEADRDVFCVADPTLRQLDHTVITVELGNIFERGINLAELPEGEQDNMGVSGA 286
DB	386 AKDEALRKQAMNGPA---VDVLTTRGLGILIRKKTIPFANLPEDIDALGISTGA 442
OY	287 GVLFGTGVMEALRTAVELFTGTPPLRLSLSEVRGMDIKETNTITMVPAGSKEELL 346
DB	443 AALFGVTGVMEALRTAVELTKGKALPKIVFEVRGLKGVRAEIDL----- 490

QY 347 KHRRAARAARAAHGTDPGLAMDGAGFTSEDRGRTILRYAVANGGLNKLITKMOAGE 406  
Db 491 -----DGR-----KIRAVVHGTAANVNLVEKTLRRE 517  
QY 407 AKDPEVIMACPGACVGGGQPRSTDKATOKROALLYNDEKSTLRSHENPSIRELD 466  
Db 518 VKHFEVMAACPGGCGGQPRSPDELIRKRAEAYITIDERLTRKSHENPAIKKYLE 577  
QY 467 TYLGEPLGKHAHELHTHYAVAGVEEKEDEK 497  
Db 578 EYLEHPLSHKHAHELHTHY-----EDSRKK 603

RESULT 2  
D57150  
hydrogenase (EC 1.18.99.1) (Fe) large chain [similarity] - Desulfovibrio fructosovorans  
C:Species: Desulfovibrio fructosovorans  
C>Date: 05-Jan-1996 #sequence\_revision 05-Jan-1996 #text\_change 19-May-2000  
C/Accession: D57150  
R:Malik, S.; Salmatalo, I.; De Luca, G.; Rousset, M.; Dermoun, Z.; Belatch, J.P.  
J. Bacteriol. 177, 2628-2636, 1995  
A:Title: Characterization of an operon encoding an NADP-reducing hydrogenase in Desulfovibrio  
A:Reference number: A57150; MUID:95270577; PMID:7751270  
A/Accession: D57150  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-585 <MAL>  
A:Cross-references: GB:U07229; NID:9466362; PION:AA87057.1; PID:9466366  
C:Superfamily: hydrogenase (Fe) large chain; ferredoxin 2(4Fe-4S) homology  
C/Keywords: 2Fe-2S; 4Fe-4S; iron; metalloprotein; oxidoreductase  
F:36/52,55,69/Binding site: 2Fe-2S cluster (Cys) (covalent) #status predicted  
F:101,105,108,111/Binding site: 4Fe-4S cluster (His, Cys, Cys, Cys) (covalent) (type N1)  
F:153,156,159,206/Binding site: 4Fe-4S cluster (Cys) (covalent) #status predicted  
F:163,196,199,202/Binding site: 4Fe-4S cluster (Cys) (covalent) #status predicted  
F:312,367,510,514/Binding site: 4Fe-4S cluster (Cys) (covalent) #status predicted  
F:514/Binding site: diron cofactor (Cys) #status predicted

Query Match 34.0%; Score 877; DB 2; Length 585;  
Best Local Similarity 41.4%; Pred. No. 4,6e-54;  
Matches 192; Conservative 66; Mismatches 138; Indels 68; Gaps 9;

QY 33 STVRVALATLEADARRLGNVA-----CAAAP-AAEAPLSHQOALAEIAPKRPDP 82  
Db 173 SGVNRGFTAVVAFAFEK-NIADTVCTNCGCVAVCPGALVEHEIYIEVEALANDP--- 228  
QY 83 TRHVCQVAVAVVAIAETGLAPGATTPKOLAEGRIGFDEVDPTLFGADLTMEEG 142  
Db 229 --KVYVQTAPAAVAAAGEDLVAPGTSYTGKMAAALRRIGFDEVDPTLFGADLTMEEG 286  
QY 143 SELLHRLTELEAHPHSDPELPMTSCCPGMIMLEKSYDPLIPVSSCKSPOMILAMV 202  
Db 287 SEFLDRIGKHLAG--DTNVKLPILTSCCPGMVAFKEHQFDFMDLDPSTASPOOMGALA 344  
QY 203 KSLIAEKKGIAAPDMVAVIMPCTRKQSEADRMFCVDADPTLRQDQDHTVTVELGNIF 262  
Db 345 KTYVADLIGIRPEKLVVSVMPCLAKKYECAARPEFSVGNP---DVDIVITTRBELAKLVK 401  
QY 263 ERGINLAELPEGEMDNPMSGAGVLFGTGVMALRTAYELPFLRLSLSEVR 322  
Db 402 RKNIDRFGRLPEDEPDAFLGASTGAAPFGVGTGVTALRTAYELAGTEFLKRVDEVDV 461  
QY 323 GMDIKETNTITWPAFGSKFEELIKHRAARAARAAHGTDPGLAMDGAGFTSEDRGCGI 382  
Db 462 GMDGVKRAKVV-----GDN 476  
QY 383 TLRYAVANGGLNKLITKMOAGEAYDEVYIMACPGACVGGGQPRSTDKATOKROA 441  
Db 477 ELVIGVANGGLNRELKPCGAGET--FHAIEVMAACPGGCGGQPRSHENPSIRELD 535  
QY 442 ALNNDEKSTLRSHENPSIRELDYTLGEPGLKHAHELHTHY 485  
Db 536 VLYAEDAGKPLKRSHPENITILEYKFLAPLSRSHOHLHTHY 579

RESULT 3  
T18557  
probable hydrogenase (EC 1.18.99.1) - Myctotherus ovalis (fragment)  
C:Species: Myctotherus ovalis  
C>Date: 02-Sep-2000 #sequence\_revision 02-Sep-2000 #text\_change 02-Sep-2000  
C/Accession: T18557  
R:Ahmady, A.; Vonken, F.; van Alen, T.; Van Hoek, A.; Boxma, B.; Vogels, G.; Veenh  
Nature 396, 527-528, 1998  
A:Title: A hydrogenosome with a genome.  
A:Reference number: Z18971; MUID:99075329; PMID:9859986  
A/Accession: T18557  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-1206 <AKX>  
A:Cross-references: EMBL:Y16775; NID:el360694; PID:el360695; PION:CA176373.1  
A:Experimental source: specific host Periplaneta americana  
C:Function:  
A:Description: catalyzes reactions involving the production or consumption of molecule  
C/Keywords: hydrogen metabolism; oxidoreductase

Query Match 32.0%; Score 824; DB 2; Length 1206;  
Best Local Similarity 42.1%; Pred. No. 6,5e-50;  
Matches 179; Conservative 72; Mismatches 120; Indels 54; Gaps 10;

QY 84 RKNVCQVAVAVVAIAETGLAPGATTPK-OLAEGRIGFDE-VPTLFGADLTME 141  
Db 241 QOQLVFOAMASIVAAEERGITRPEKILNEATATRLKGSNVFLDTNFSADLTITE 300  
QY 142 GSELHRLTEH-----LEAHPHSDPELPMTSCCPGMIMLEKSYDPLIPVSSCKSPOM 196  
Db 301 GHLEILRYLVNNGKMLLGDDHNPDLPMLTSCCPGMIMLEKSYDPLIPVSSCKSPOM 360  
QY 197 MLAARVSYIAEK-KGIAPDMVAVIMPCTRKQSEADRMFCVDADPTLRQDQDHTVTV 255  
Db 361 MGALIKGVYAKKIKKDPDIYSVIMPCTAKKAEKERO--LRGDEGKVDYITLTR 418  
QY 256 ELGINFERGINLAELPEGEMDNPMSGAGVLFGTGVMALRTAYELFTG--TPL 313  
Db 419 ELAKMLKQSNIDIAKEPPFDVYMESEGTAAVIFGTGVMALRTAYELFTG--TPL 478  
QY 314 PRLSLEVRGMDIKETNTITWPAFGSKFEEL-KHRAARAARAAHGTDPGLAMDGAG 372  
Db 479 KNLNIEAVRGMGIREA-----GKLENVLDYKAFB----- 510  
QY 373 FTSEDRGRTILRYAVANGGLNKLITKMOAGEAYDEVYIMACPGACVGGGQ 427  
Db 511 -----GTVAVAAIAHGNNAKRVMDIIRKQKESGKPAHFEVMAACPGGCGGQ 562  
QY 428 PRSTDKATOKROALLYNDEKSTLRSHENPSIRELDYTLGEPGLKHAHELHTHY 487  
Db 563 PKPTNLEIRAKRQTLPTKEMDPLRKSHNPETIKATYENLKEPLGHNHHLHTYSS 622  
QY 488 GGVEE 492  
Db 623 OKVRD 627

RESULT 4  
S13526  
hydrogenase (EC 1.18.99.1) (Fe) large chain [similarity] - Desulfovibrio vulgaris  
N:Alternate names: hydrogenase gamma  
C:Species: Desulfovibrio vulgaris  
C>Date: 23-Apr-1993 #sequence\_revision 23-Apr-1993 #text\_change 20-Apr-2001  
C/Accession: S13526  
R:Stokkermans, J.; van Dongen, W.; Kaan, A.; van den Berg, W.; Veeger, C.  
FEBS Microbiol. Lett. 49, 217-222, 1988  
A:Title: hyd-gamma, a gene from Desulfovibrio vulgaris (Hildenborough) encodes a poly  
A:Reference number: S13526; MUID:89306533; PMID:2663654  
A/Accession: S13526  
A:Molecule type: DNA





Db 182 WOKVATYYPELLPHFSTCKSPIGNGALAKTYGAEKRMKYPDKOYTVSIMPICAKRYEG 241  
OY 233 DEDMFCVADPTL-----RDLHVITTVTELGNIFERGINLAELPEGEMDNMGVSGAG 287  
Db 242 LR-----PELKSSGRDIDATITTRRLAYMIKKAGIDGFKLPDGKDSLMGESTGA 293  
OY 288 VLEFGTGGVMEALRTAYELFTGTPLPRLSLSEVRGMGDIKETNITMVPAPSGKPEELK 347  
Db 294 TIFGVTGGVMEALRFAVEAYVTKKPPDSMDPKAVGGLDGIKEATVNV----- 340  
OY 348 HRAAARAEAAHGTGCPPLAMDGAGFTSDEGRGITLRYAVANGCNAKKLITKMOGAE 407  
Db 341 -----GCTDVKAYAVHGAKKRFQVCDDVYAGKS 368  
OY 408 KYDFEIMACPGCVGGGQP-----RSTDKAIRQ-----KROAL 443  
Db 369 PHELEIMACPGCGVGGGQPVMPGVLEMDRTTRRLVAGLKKRLAM 415

RESULT 9  
HODVIV  
Cytchrome-c3 hydrogenase (EC 1.12.2.1) (Fe) large chain - Desulfovibrio vulgaris subsp.  
N:Alternate names: hydrogenase (Fe) alpha chain  
C:Species: Desulfovibrio vulgaris subsp. oxamunicus  
C:Date: 31-Dec-1990 #sequence\_revision 31-Dec-1990 #text\_change 05-May-2000  
C:Accession: A32886  
R:Voordouw, G.; Strang, J.D.; Wilson, F.R.  
J. Bacteriol. 171, 3881-3889, 1989  
A:Title: Organization of the genes encoding [Fe] hydrogenase in Desulfovibrio vulgaris  
A:Reference number: A32886; MUID:89291738; PMID:2661538  
A:Accession: A32886  
A:Molecule type: DNA  
A:Residues: 1-421 <VOO>  
C:Superfamily: hydrogenase (Fe) large chain; ferredoxin 2[4Fe-4S] homology  
C:Keywords: 4Fe-4S; hydrogen metabolism; iron-sulfur protein; metalloprotein; oxidoreduc  
F:28-84/Domain: ferredoxin 2[4Fe-4S] homology <FE>  
F:35-38/41/6/Binding site: 4Fe-4S cluster (Cys) (covalent) #status predicted  
F:45-66/69/72/Binding site: 4Fe-4S cluster (Cys) (covalent) #status predicted  
F:382/Binding site: diron cofactor (Cys) #status predicted  
C:Genetics:  
A:Gene: hvdA  
C:Superfamily: hydrogenase (Fe) large chain; ferredoxin 2[4Fe-4S] homology  
C:Keywords: 4Fe-4S; hydrogen metabolism; iron-sulfur protein; metalloprotein; oxidoreduc  
F:28-84/Domain: ferredoxin 2[4Fe-4S] homology <FE>  
F:35-38/41/6/Binding site: 4Fe-4S cluster (Cys) (covalent) #status predicted  
F:45-66/69/72/Binding site: 4Fe-4S cluster (Cys) (covalent) #status predicted  
F:382/Binding site: diron cofactor (Cys) #status predicted

Query Match 23.3%; Score 599.5; DB 1; Length 421;  
Best Local Similarity 36.5%; Pred. No. 1.1e-34;  
Matches 135; Conservative 49; Mismatches 117; Indels 69; Gaps 5;

OY 87 VCVQVAPAVRAVAIAETLGLAPGATPPKQALAEGLRGLGDEVDLTGADLTMEBSSELL 146  
Db 102 VAMPAPAVRAVAIAEGLFVGTVTGKMFSAKELGFDHCDNDEFTADVTIWEESTEFY 161  
OY 147 HLTETLEAHPSDEPLPMTSCCPGMIMLEKSYDLPYVSSCKSPOMMLAAMVSKYL 206  
Db 162 QRLTKL-----DKPLPQFTSCCPGMIMVLESILPELPHMSSCSPIGMJSTLTKTIG 215  
OY 207 AEKGIAPKDMVMSIMPTCKROSEADRP--WFCVADADPTLRQLDHLVITTVTELGNIFER 264  
Db 216 ADRMKYDRAKYTVTSIMPTAKRYEGRPOLW-----DSGHKDIDITITRRLAYMIKKA 270  
OY 265 GINTLAELPGEKDNPMGVSGAGVLTFTTGGVMEALRTAYELFTGTPLPRLSLSEVRGM 324  
Db 271 KIDFLKLPDGKSDTLMEGSTGATLFGVTGGVMEALRTAYQAVTVTKKSPESMDPKFVRCI 330  
OY 325 DKIKETNITMVPAPSGKPEELKXRAAARAEAAHGTGCPPLAMDGAGFTSDEGRGITL 384  
Db 331 QGVKEKTVNV-----GGVDV 345  
OY 385 RVAVANGCNAKKLITKMOGAEKAYDFEIMACPGCVGGGQP-----RSTDK 433

Db 346 KVAVHAGARRFHDVCELVAKRAPHFIEFMACPGCGVGGGQPVMPGVLEADRRSTEM 405  
OY 434 AITOKROAL 443  
Db 406 YAGLKKRLAM 415

RESULT 10  
G72256  
hydrogenase (EC 1.18.99.1) (Fe) large chain [similarity]-- Thermotoga maritima (stra  
C:Species: Thermotoga maritima  
C:Date: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 21-Jul-2000  
C:Accession: G72256  
R:Nelson, K.E.; Clayton, R.A.; Gail, S.R.; Gwinn, M.L.; Dodson, R.J.; Hafey, D.H.; Hic  
Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson,  
C.M.  
Nature 399, 323-329, 1999  
A:Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome  
A:Reference number: A72200; MUID:99287316; PMID:10360571  
A:Accession: G72256  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-645 <ARN>  
A:Cross-references: GB:AE001794; GB:AE000512; NID:94981978; PIDN:AAJ36496.1; PID:9498  
A:Experimental source: strain MSB  
C:Genetics:  
A:Gene: TM1426  
C:Superfamily: hydrogenase (Fe) large chain; ferredoxin 2[4Fe-4S] homology  
C:Keywords: 2Fe-2S; 4Fe-4S; iron; metalloprotein; oxidoreductase  
F:34/45/48/60/Binding site: 2Fe-2S cluster (Cys) (covalent) #status predicted  
F:92/96/99/105/Binding site: 4Fe-4S cluster (His, Cys, Cys, Cys) (covalent) (type N1)  
F:143/146/149/196/Binding site: 4Fe-4S cluster (Cys) (covalent) #status predicted  
F:153/186/189/192/Binding site: 4Fe-4S cluster (Cys) (covalent) #status predicted  
F:293/350/482/486/Binding site: 4Fe-4S cluster (Cys) (covalent) #status predicted  
F:486/Binding site: diron cofactor (Cys) #status predicted

Query Match 19.9%; Score 512.5; DB 2; Length 645;  
Best Local Similarity 30.2%; Pred. No. 2.6e-28;  
Matches 133; Conservative 56; Mismatches 159; Indels 93; Gaps 9;

OY 54 CAAPAPABAPLSHVQALAEKPRDDPTKRVQVAVRAVAIAETLGLAPGATTPK 113  
Db 192 CVAVCPGALSTINDIKILEALE-----SDKIVIGMIPAVRAAIOEERGIODEVMAE 246  
OY 114 QLAEGIRLGFDEVDLTGADLTMEBSSELLRTEHLEAHPSHDEPLPMTSCCPGM 173  
Db 247 KIVSFLTKTIGFDKVFVDFVSGADLVAYEEAHEFYERLKK-----GERLPQFTSCCPAW 298  
OY 174 IAWLEKSYDLPYVSSCKSPOMMLAAMVSKYLAERKGLAPKDMVMSIMPTCKROSEAD 233  
Db 299 VKHAERTYPOYLONLSSVSPQALGTIVIKIARKLGYVEEKLPLVSPFCTAKKFEAL 358  
OY 234 RMFCVADPTLRQLD-----VITTVELGNIFERGINLAELPEGEMDNPMGVSGAG 287  
Db 359 RE-----EHEGIVDIVLFTRELALQIKSRIDINVERPOFPDRPYGVSQAG 405  
OY 288 VLEFGTGGVMEALRTAYELFTGTPLPRLSLSEY---RGMGDIKETNITMVPAPSGKFE 344  
Db 406 LGFGKAGVGSCLVSLNE-----EIGIEIVDVKSPEBDGIRVABVTL----- 447  
OY 345 LKXRAAARAEAAHGTGCPPLAMDGAGFTSDEGRGITLRYAVANGCNAKKLITKMOA 404  
Db 448 -----KQTSFGKAVIYIGLGVKKFLERK- 472  
OY 405 GEAKYDFEIMACPGCVGGGQPRSTDRATITOKROALYMLDEKSTLRSHENPSITREL 464  
Db 473 ---DVEIIEVMACNCGVGGGQPYPDNSRIREHRAKVLADTGIRKSLTLTPVENLFMLKL 529  
OY 465 YDTYIGEPRLGKRAHEHLHTHY 485  
Db 530 YEEDLKDE--HTRHEILHTTY 548

```

RESULT 11
T40992
hyothetical protein SPCC1450.10c - fission yeast (Schizosaccharomyces pombe)
C:Species: Schizosaccharomyces pombe
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
C:Accession: T40992
R:Lynne, M.; Rajandream, M.A.; Barrell, B.G.; Volckaert, G.
submitted to the EMBL Data Library, March 1999
A:Reference number: 221962
A:Accession: T40992
A:Status: preliminary; translated from GB/EMBL/DDJ
A:Molecule type: DNA
A:Residues: 1-538 <L>N>
A:Cross-references: EMBL:AL049559; PIDN:CA940177.1; GSPDB:GN00068; SPDB:SPCC1450.10c
A:Experimental source: strain 972H; cosmid c1450
C:Genetics:
A:Gene: SPDB:SPCC1450.10c
A:Map position: 3
A:Introns: 31/3

Query Match
Best Local Similarity 24.1%; Score 311.5; DB 2; Length 538;
Matches 123; Conservative 84; Mismatches 179; Indels 123; Gaps 19;

OY 52 VACAAAAPAAEAPLASHVQALAEALAPKDDPTKRWCVQYAPAVAAIAETGLACGATT 111
DB 65 LACSGCITSSEITLVNLSQYQEVKLHLESRSQELIYVSLSPQVRAALAAVYGLSLQEIQ 124
OY 112 PKLAGSLRLGFDEVDPTLFGADLTIMESGSELRLHLEHAAH----- 156
DB 125 AVLEMPFICKGLGFAHLDITNAREIYLOCAQCEFCNSWLOS-RHKNQNOVTSVYNEHP 183
OY 157 --PHS-----DEPLMFTSCCPGMIALEKSYPDILPYVSSCKSPQ 195
DB 184 LIPHSSTQISGVHSNTSSNGINENAVLPILSSSCPGMICYVEKTHSNLIPNLSRVASFPQ 243
OY 196 MMLAAVKSILAEEKGIAPADWVWSIMPTCKQSEADRMFCVADPTLRDLHVTYTV 255
DB 244 QACGRILKDWAVOQFSQNRDVMHLSLMPCTFKLEASRDEF--SENGVADSVLTTPK 300
OY 256 ELGINIFERGINLAELPEGEWMDPM-----GVSAGVLFGTGG 295
DB 301 ELVEMKFLRIDIELETK---NPFPQOSTDAIPFYPRIITYEBOIGS-----SSGG 349
OY 296 VMEALATAYELFTGTP--LPRLSLSEVRGMDGIK-----ETN---ITVVPAPGSKFEE 344
DB 350 YMGVYLSYAAKMLFEGIDDPYVSMNKNK-DLEETYLRLHPETNEGLISMACTCYGFRNIQ 408
OY 345 LKHHRAAARAEAAHGTGPPGLANDGAGFTSEDRGSGITLRVAVANGLCNAKKLITKMQA 404
DB 409 NLVRR-----VHG-----NSSVAKRGVLLKKRVRSNAQNPTE----- 440
OY 405 GEAKYDFVEIMACPGACVGGGG--PRSTDKAITQKRQ-----AALYMLDKSTLRSHEN 458
DB 441 EPRYDVEVMACPGGCGINGGOLPEPVSERIVASARMMQOEVETLY--EPGT--RSVDQ 496
OY 459 PSIRELYDTYLGEPLGKAHLLHTHYVA 487
DB 497 SAVSYMLEQWVKDPT--LTPKFLHTSYRA 523

RESULT 12
S63206
Left protein homolog YNL240c - yeast (Saccharomyces cerevisiae)
N:Alternate names: hypothetical protein N1114
C:Species: Saccharomyces cerevisiae
C:Date: 27-Apr-1996 #sequence_revision 03-May-1996 #text_change 19-Apr-2002
C:Accession: S63206; J00891; S67358; S72076
R:Pandolfo, D.; De Antoni, A.; Lanfranchi, G.; Valle, G.
submitted to the Protein Sequence Database, April 1996
A:Reference number: S63206
A:Accession: S63206
A:Molecule type: DNA

A:Residues: 1-491 <P>N>
A:Cross-references: EMBL:271516; NID:g1302273; PID:e239709; PID:g1302274; MIPS:YNL240
A:Experimental source: strain S288C
R:Nogee, I.; Johnston, M.
Gene 96, 161-169, 1990
A:Title: Isolation and characterization of the ZNF1 gene of Saccharomyces cerevisiae,
A:Reference number: J00890; MID:91095671; PMID:2269430
A:Accession: J00891
A:Molecule type: DNA
A:Residues: 229-280, 'AIRA', 286-374, 'SSAE', 380-392, 'P', 396-463, 'PPA', 473-474, 'R', 476,
A:Cross-references: GB:M34709; NID:g171543; PIDN:AAA34618.1; PID:g171544
R:Pandolfo, D.; de Antoni, A.; Lanfranchi, G.; Valle, G.
submitted to the EMBL Data Library, February 1996
A:Description: DNA sequence of cosmid 14-5 from chromosome XIV.
A:Reference number: S67355
A:Accession: S67358
A:Molecule type: DNA
A:Residues: 1-491 <P>N>
A:Cross-references: EMBL:269381; NID:g1183970; PID:e221952; PID:g1183974
R:Pandolfo, D.; de Antoni, A.; Lanfranchi, G.; Valle, G.
Yeast 12, 1071-1076, 1996
A:Title: The DNA sequence of cosmid 14-5 from chromosome XIV reveals 21 open reading
A:Reference number: S72073; MID:97051596; PMID:8896273
A:Accession: S72076
A:Status: nucleic acid sequence not shown; translation not shown
A:Residues: 1-491 <P>N>
A:Molecule type: DNA
A:Cross-references: EMBL:269381; NID:g1183970; PIDN:CAA93358.1; PID:g1183974
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, February 1996
C:Genetics:
A:Gene: SGD:NA1
A:Cross-references: SGD:S0005184
A:Map position: 14L
A:Note: YNL240c

Query Match
Best Local Similarity 23.1%; Score 255; DB 2; Length 491;
Matches 114; Conservative 72; Mismatches 189; Indels 118; Gaps 18;

OY 52 VACAAAAPAAEAPLASHVQ-----QALAEALAPKDDPTKRWCVQYAPAVAAIAETGL 105
DB 60 LACSGCITSSEITLVNLSQSHSVFLKMKNGKLSQGD-----KFLVSVSPQRLSLAQYGL 115
OY 106 APGATPPKOLAEGRLGFDEVDPTLFGADLTME-----EGSEL-----LHRL 149
DB 116 -----TEAADICLNNFQKHFQCKYWGTEMGRTISISKY 151
OY 150 TEHLEAHPHSDPL-----PMFTSCCPGMIALEKSYPDILPYVSSCKSPQMLAAVKS 204
DB 152 VEKTIHKKOKENTGADRKLPLSAVCPGFLYTEKTPQVLPMILNKSPOQITGSLIRA 211
OY 205 YLAEEKGIAPKDWVWSIMPTCKQSEADRMFCVADPTLRDLHVTYTVELGINIFKER 264
DB 212 TPF-ESLAIARESYHLSLMCFDKLEASRDELDS-----IDCYITPREIYTMQEL 264
OY 265 GINLAEL-----PEGEWMDPMVGSAGVLFGTGGVMEALATAYELFTGT- 311
DB 265 NLDFKSFUTBDTSLYGRSPG--WDPRVHASNIG--GTCGGY-----AYQYTVAYQ 313
OY 312 ---PLPRLSLSEVRGMDGIGETNITVVPAPGSKFEELTKR---AAARAEAAHGTGPP 365
DB 314 RLHPGSMIVLEGNSDIV-----EYRLLDHRIIAASLSGFRNIQNL 358
OY 366 AMDGAGFTSEDRGSGITLRVAVANGLCNAKKLITKMQAGEARY--DFVEIMACPGACV 423
DB 359 VKLTSGSGSEKRNITLALPKRRTGPRANSREMAAATAADYHSDIYEVNCPGCMN 418
OY 424 GGGPSTDKAITQKQALYLNDEKSTLRSHENPSIRELYDTYLGEPLGKAHLLHT 483
DB 419 GCGILNENQNSLKRKQLV-----QTLNKRH--GEELAVVDLTGLPKLEAAARPLSL 469
OY 484 HYVAGVYEKDEK 496

```

Db 470 EYFAPVQAVK 482

## RESULT 13

hydrogenase chain (ferredoxin) [imported] - Clostridium acetobutylicum

C:Species: Clostridium acetobutylicum

C:Date: 14-Sep-2001 #sequence\_revision 14-Sep-2001 #text\_change 14-Sep-2001

C:Accession: B97297

R:Rolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee, J.; Daily, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.

J. Bacteriol. 183, 4823-4838, 2001

A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Clostridium acetobutylicum

A:Reference number: A96900; MUID:21359325; PMID:13593325

A:Accession: B97297

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-450 &lt;KOR&gt;

A:Cross-references: GB:AE001437; PIDN:AAK81165.1; PID:q15026303; GSPDB:GN00168

A:Experimental source: Clostridium acetobutylicum ATCC824

C:Genetics:

A:Gene: CAC3230

Query Match 9.5%; Score 245.5; DB 2; Length 450; Best Local Similarity 24.6%; Pred. No. 1.1e-09; Matches 96; Conservative 50; Mismatches 113; Indels 131; Gaps 14;

QY 87 VCQVAPAVRAIAETLGLAATTPKQALBGLRLGDEVDITLFGAD-LTME--ES 143

DB 155 VTAAPAPAIAGQFGENVSL-----EMLRFAFKKVGADNVEAFADMLTIKAEFEEN 207

QY 144 ELHLRLTELEAHPSDEPLPMFTSCCPGMIMLEKSPDLPYVSSCKSPOMLAAMVK 203

DB 208 ELV-----NSKQDLMTTSCCCPMWVMIRKIKYDLKRAHSPSPVPIASGRYIK 256

QY 204 SYLAEKGIAPKDMVMSIMPTKROSEADRDVCFADPTLRQDLHYITVELGNITKE 263

DB 257 KLNPMCK-----VVEFTGPIAKKAS-RSODISDA-----IDFVLTFELGIFDV 301

QY 264 RGINALAEPLPEGMNDPMGVSGAVLFTGTGVM---EALR---TAYELTGPIPLPL 316

DB 302 LDIDELKPLPE--THTKSYASREGLYGTGVSTVDEAVYRIPPNKHILFKSTK--- 354

QY 317 SLSEVGMGDIKETNITWVPAPGSKFEELKHRAARAFAAHGTPGLAMDGAGFTSE 376

DB 355 -----VDGVKD----- 360

QY 377 DGRGITLRVAVANGIAGNAKKLITKMGAGEAKYDFVEIMACPAGCGGGGPRSTDKAIT 436

DB 361 -----CKDILNKTQAGNIGANFLEGCGVCGCGV-----PKATV 395

QY 437 QKROA--ALXNDEKSTLRSHENSIREL 464

DB 396 HKDGRESVNMKTAESEIKTIVSDSRMKDI 425

## RESULT 14

S31336

LEFT1 protein - yeast (Kluyveromyces marxianus var. lactis)

C:Species: Kluyveromyces marxianus var. lactis, Candida sphaerica

C:Date: 13-Jan-1995 #sequence\_revision 13-Jan-1995 #text\_change 29-Oct-1999

C:Accession: S31336

R:Mesolowski-Louvet, M.; Tanguy-Rougeau, C.; Fuhkhan, H.

submitted to the EMBL Data Library, February 1993

A:Reference number: S31336

A:Accession: S31336

A:Molecule type: DNA

A:Residues: 1-469 &lt;MES&gt;

A:Cross-references: EMBL:X70373; NID:95537; PIDN:CAA49833.1; PID:95538

C:Genetics:

A:Gene: LEFT1

Query Match 9.0%; Score 232; DB 2; Length 469;

Best Local Similarity 23.9%; Pred. No. 1e-08; Matches 110; Conservative 62; Mismatches 167; Indels 122; Gaps 16;

QY 52 VACAAAPAAEAPLSHVQOALAEAK--PKDDPTKHCVCVAPAVRAIAETLGLAPGA 109

DB 57 IACACITSSSEILLSKOSHVKVLEKMSLELEIDERSLAVSISPQRLSLADYSC--- 113

QY 110 TTPKQALBELR-----RLGFDEVFDLRFADLTIMESSELHLRLTEHLEAHPSDE 161

DB 114 -----LADIDRCFQNFMTKFNKAYVGVGFGRSISIRNATLDRPEN----- 159

QY 162 PLPMFTSCCPGMIMLEKSPDLPYVSSCKSPOMLAAMVSYLAEEKGIAPKDMVMS 221

DB 160 BEPFLCSVCPGVLAIEAKRPFLIHMIDYKSPQDTGLMLKQ-----ADPCYHLS 211

QY 222 IMPCTRKQSEADRDVCFVADADPTLRQDLHYITVELGNITKERNL-----ABL 271

DB 212 IMPCFDKKLEASR-----EECEKEVDVIRPKQFVAMLGDSIDFKSYMTEDYSKEL 264

QY 272 PEGEMDNPBGVSGAVLFTGTGVMALR-----TAYELFG 310

DB 265 CPSSGMDYKILMLSNBQ---SSSGYAYOTLSLQSNPESDIITTEGKNSDVTETRLVSK 321

QY 311 TPLPRLSLSEVRGMDGI-----KETNITWVPAPGSKFEELKHRAARAFAAHGTPPL 365

DB 322 SKGVLASSEVYGFNRIOVLKRLQSASVKKRGKIK---YARRGSVYK----- 367

QY 366 AMDGAGFTSEDCRGITLRVAVANGIAGNAKKLITKMGAGEAKYDFVEIMACPAGCGG 425

DB 368 -----SGEISE-----KTRKYLTL---ADPAKTFVEVYMACPSCGIC 402

QY 426 GQPRSTDKAITOKROALYNLDEKSTLRSHENSIRELYD 466

DB 403 GL-LNEKMANRRKOLA-ODLSLAYTKVHSVNPIDVIAVD 441

## RESULT 15

HODVFS

hydrogenase (EC 1.18.99.1) (Fe) small chain precursor - Desulfovibrio vulgaris (strain

Naltermale names: Fe hydrogenase beta chain

C:Species: Desulfovibrio vulgaris

C:Date: 31-Dec-1990 #sequence\_revision 31-Dec-1990 #text\_change 16-Jun-2000

C:Accession: B24551; A27480

R:Voordouw, G.; Brenner, S.

Eur. J. Biochem. 148, 515-520, 1985

A:Title: Nucleotide sequence of the gene encoding the hydrogenase from Desulfovibrio

A:Reference number: A24551; MUID:85203856; PMID:3888621

A:Accession: B24551

A:Molecule type: DNA

A:Residues: 1-123 &lt;VOO&gt;

A:Cross-references: GB:X02416; NID:940829; PIDN:CAA26267.1; PID:9440831

R:Prickall, B.C.; He, S.H.; Li, C.; Menon, N.; Choi, E.S.; Przybyla, A.E.; DerVartanli

Biochem. Biophys. Res. Commun. 149, 369-377, 1987

A:Title: Identification of three classes of hydrogenase in the genus, Desulfovibrio.

A:Reference number: A27480; MUID:88106446; PMID:3322725

A:Accession: A27480

A:Status: preliminary

A:Molecule type: protein

A:Residues: 35-69 &lt;PRI&gt;

C:Comment: Three distinctive types of hydrogenases, the Fe, NiFe, and NiFeSe, are found

C:Comment: This hydrogenase complex contains three (Fe-4S) iron-sulfur clusters.

C:Genetics:

A:Gene: hvdB

C:Complex: heterodimer of large (alpha, see PIR:HODVFL) and small (beta) chains

A:Description: catalyze reactions involving the production or consumption of molecular

A&gt;Note: may be involved in hydrogen uptake for the reduction of sulfate to hydrogen s

C:Superfamily: hydrogenase (Fe) small chain

C:Keywords: hydrogen metabolism; iron-sulfur protein; oxidoreductase; periplasmic spa

F1-34/Domain: signal sequence (status predicted &lt;SD&gt;

F:35-123/Product: hydrogenase (Fe) small chain (status predicted &lt;MAT&gt;

Query Match 5.4%; Score 138.5; DB 1; Length 123;

Best Local Similarity 38.7%; Pred. No. 0.0068;  
Matches 29; Conservative 14; Mismatches 21; Indels 11; Gaps 2;

```

QY      429 RSTDKATIQKROA-----ALYNLEKSTLRSHENPSTRELYDTYGEPLGKAHE 479
        | | ||: | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      27 RWTGAAVAAYQIKDYMLDRINGVYGAADKFPVRASQDNTQYKALYKSTYKPLGKSHD 86

```

Db 27 RMTGKAVAAVKQIKDYMLDRINGVYGADAKFPVRASQDNTQVKALYKSYLEKPLGHKSHD 86

QY 480 LHTHYV--AGGVEE 492

Db 87 LHTHWFDKSKVKE 101

Search completed: June 3, 2003, 16:17:10  
Job time : 45 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 3, 2003, 16:07:42 ; Search time 22 Seconds

(without alignments)  
936.988 Million cell updates/sec

Title: CAC80065

Perfect score: 2576

Sequence: 1 MSALYKPCAAVNSINGSCR.....HELTHTHYAGVEEKDEKK 497

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_40.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	743.5	28.9	574	1 PHFL_CLOPA	P29166 Clostridium
2	633	24.6	421	1 PHFL_DESVH	P07598 desulfofibr
3	599.5	23.3	421	1 PHFL_DESVO	P13629 desulfofibr
4	255	9.9	491	1 NARL_YEAST	P23603 saccharomyc
5	232	9.0	469	1 LETI_KLULA	P53998 kluyveromyc
6	138.5	5.4	123	1 PHFS_DESVH	P07603 desulfofibr
7	130	5.0	124	1 PHFS_DESVO	P13628 desulfofibr
8	112	4.3	776	1 HYPE_AZOVI	P40596 azotobacter
9	111	4.3	912	1 ICAS_RABIT	Q28730 cryotolagus
10	109.5	4.2	857	1 CLPB_ECOLI	P03815 escherichia
11	108.5	4.2	1887	1 PAS2_YEAST	P19097 s fatty aci
12	107.5	4.2	391	1 ALR_STRCO	O86786 streptomyc
13	106	4.1	1885	1 PAS2_CANAL	P43098 c fatty aci
14	105.5	4.1	3491	1 ERYL_SACER	P03131 saccharopol
15	105	4.1	447	1 ODP2_RHIME	Q91903 rhizobium m
16	105	4.1	896	1 IF2_CHLMU	Q98K00 chlamydia m
17	104	4.0	391	1 THIL_PARDE	P54810 paracoccus
18	103.5	4.0	377	1 PROB_METRU	Q92998 meliothermus
19	103.5	4.0	1799	1 LMB2_MOUSE	O61592 mus musculu
20	102	4.0	835	1 RIRL_PPVKA	P50643 pseudotable
21	100	3.9	1842	1 PAS2_SCHPO	O10289 s fatty aci
22	99	3.8	479	1 DP04_STRCO	Q94882 streptomyc
23	99	3.8	1322	1 PUR4_XYLFA	Q94876 xylaria fas
24	98.5	3.8	395	1 PURK_STNP7	Q54975 synchococc
25	98.5	3.8	871	1 ARS6_NEUCR	P54998 n arg-6 pro
26	98.5	3.8	3695	1 LMA5_HUMAN	O15330 homo sapien
27	97	3.8	388	1 ALR2_AGR75	P58737 agrobacteri
28	97	3.8	454	1 UCR2_NEUCR	O60044 neurospora
29	96.5	3.7	732	1 YPA8_HUMAN	Q9hcm4 homo sapien
30	96	3.7	1181	1 PIR2_HUMAN	O00722 homo sapien
31	95.5	3.7	553	1 SPK_RAT	O88506 rattus norv
32	95.5	3.7	686	1 EFG2_STRCO	O87644 streptomyc
33	95.5	3.7	810	1 SYFB_SYNY3	P74296 synchocyst

34	95	3.7	247	1 ATFD_SORBI	O07300 sorghum bic
35	95	3.7	511	1 DLOH_SCHPO	O00087 schizosacch
36	95	3.7	1857	1 PAS2_PENPA	P15368 p fatty aci
37	95	3.7	2541	1 TAL1_HUMAN	Q91490 homo sapien
38	95	3.7	2541	1 TAL1_MOUSE	P26039 mus musculu
39	94.5	3.7	971	1 KBFL_MOUSE	P25799 mus musculu
40	94	3.6	748	1 CLAT_HUMAN	P28329 homo sapien
41	94	3.6	1487	1 ICP4_HSEK	P17473 equine hepr
42	94	3.6	2035	1 HPCL_HUMAN	P51610 homo sapien
43	93.5	3.6	394	1 AT0B_ECOLI	P76461 escherichia
44	93	3.6	563	1 ARAB_BACHD	Q9Kbq3 bacillus ha
45	93	3.6	673	1 VID3_AGR75	P18593 agrobacteri

## ALIGNMENTS

```

RESULT 1
PHFL_CLOPA
ID PHFL_CLOPA STANDARD: PRT: 574 AA.
AC P29166;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Periplasmic [Fe] hydrogenase 1 (EC 1.18.99.1).
OS Clostridium pasteurianum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1501;
ON [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC STRAIN=ATCC 6013 / W5.
RX MEDLINE=92002005; PubMed=1911757;
RA Meyer J., Gagnon J.;
RT Primary structure of hydrogenase I from Clostridium pasteurianum.;
RL Biochemistry 30:9697-9704(1991).
RN [2]
RP X-RAY CRYSTALLOGRAPHY (1.8 ANGSTROMS).
RX MEDLINE=99055388; PubMed=9836629;
RA Peters J.W., Lanzillo W.N., Lemon B.J., Seefeldt L.C.;
RT X-ray crystal structure of the Fe-only hydrogenase (Cp1) from
RL Clostridium pasteurianum to 1.8-A resolution.;
RC Science 282:1853-1858(1998).
CC -1- CATALYTIC ACTIVITY: 2 reduced ferredoxin + 2 H(+) -> 2 oxidized
CC ferredoxin + H(2)
CC -1- COPROCTOR: THIS PROTEIN BINDS PROBABLY FOUR AFE-4S CLUSTERS AND ONE
CC H CLUSTER (POSSIBLE SIX-TO-EIGHT IRON CLUSTER).
CC -1- SUBUNIT: MONOMER.
CC -1- SUBCELLULAR LOCATION: Periplasmic.
CC -1- SIMILARITY: THE IRON-SULFUR CENTERS ARE SIMILAR TO THOSE OF
CC "BACTERIAL-TYPE" AFE-4S FERREDOXINS.
CC -1- SIMILARITY: TO THE D-VULGARIS (HIDENBROUGH) HYD-GAMMA GENE
CC PRODUCT.
-----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL: M61737; AAA23248.1; ..
CC PIR: A40330; HOCLIP.
CC PDB: 1FEH; 06-JAN-99.
CC InterPro: IPR001450; 4Fe4S ferredoxin.
CC InterPro: IPR003149; Fe_hyd_SSU.
CC InterPro: IPR004108; Fe_hyd_1g_C.
CC InterPro: IPR001041; Ferredoxin.
CC Pfam: PF00037; fer4; 2.
CC Pfam: PF00111; fer2; 1.
CC Pfam: PF02256; Fe_hyd_SSU; 1.
CC Pfam: PF02906; Fe_hyd_1g_C; 1.

```

DR PROSITE; P500198; 4FE4S\_FERREDOXIN; 2.  
 KW Oxidoreductase; Periplasmic; Iron-sulfur (4FE-4S) (POTENTIAL).  
 FT METAL 34 34 IRON-SULFUR (4FE-4S) (POTENTIAL).  
 FT METAL 46 46 IRON-SULFUR (4FE-4S) (POTENTIAL).  
 FT METAL 49 49 IRON-SULFUR (4FE-4S) (POTENTIAL).  
 FT METAL 62 62 IRON-SULFUR (4FE-4S) (POTENTIAL).  
 FT METAL 98 98 IRON-SULFUR (4FE-4S) (POTENTIAL).  
 FT METAL 101 101 IRON-SULFUR (4FE-4S) (POTENTIAL).  
 FT METAL 107 107 IRON-SULFUR (4FE-4S) (POTENTIAL).  
 FT METAL 147 147 IRON-SULFUR 3 (4FE-4S) (BY SIMILARITY).  
 FT METAL 150 150 IRON-SULFUR 3 (4FE-4S) (BY SIMILARITY).  
 FT METAL 153 153 IRON-SULFUR 3 (4FE-4S) (BY SIMILARITY).  
 FT METAL 157 157 IRON-SULFUR 3 (4FE-4S) (BY SIMILARITY).  
 FT METAL 190 190 IRON-SULFUR 4 (4FE-4S) (BY SIMILARITY).  
 FT METAL 193 193 IRON-SULFUR 4 (4FE-4S) (BY SIMILARITY).  
 FT METAL 196 196 IRON-SULFUR 4 (4FE-4S) (BY SIMILARITY).  
 FT METAL 200 200 IRON-SULFUR 4 (4FE-4S) (BY SIMILARITY).  
 FT METAL 299 299 IRON-SULFUR (H CLUSTER) (POTENTIAL).  
 FT METAL 300 300 IRON-SULFUR (H CLUSTER) (POTENTIAL).  
 FT METAL 355 355 IRON-SULFUR (H CLUSTER) (POTENTIAL).  
 FT METAL 499 499 IRON-SULFUR (H CLUSTER) (POTENTIAL).  
 FT METAL 503 503 IRON-SULFUR (H CLUSTER) (POTENTIAL).  
 SQ SEQUENCE 574 AA; 63828 MW; 17E28A74E23C7DEE CRC64;  
 Query Match 28.9%; Score 743.5; DB 1; Length 574;  
 Best Local Similarity 37.1%; Pred. No. 2.1e-44;  
 Matches 163; Conservative 65; Mismatches 142; Indels 69; Gaps 6;  
 QY 52 VCAAAAPAAEAPLSHVOQALAEIAKPKDDPRKHYCVQVAPAVRAIAETLGLAPGATT 111  
 DB 198 IACPVAAAEKSHMDRVKAL-----NAPEKHVYAAAPSVRAISIELEFNMGVDV 249  
 QY 112 PROAEGRLRLGFDEVDTLFGADLTITMEGSELLRLTEHLEAPHSDEPLPMFTSCPP 171  
 DB 250 TGRITATLQGLQDFDKFDLNFADMTIMEATELVORI-----ENNPPPMFTSCPP 301  
 QY 172 GRIAMLEKSPDLIPYVSSCKSPOMLAAMVKSYLEAKGAPKDMVAVSIMPCTRKOSE 231  
 DB 302 GAVRAOENYYPRLINMISAKSPQFGTASTYYSISGLDPKKNFTVTVMCTSKPFE 361  
 QY 232 ADRDMCVADPTLRDLHIVITTVELGNIFKRGCIILAEIPSEMDNPVGSAGAVLGG 291  
 DB 362 ADRPOMEKCG---LRIDAVITTRELAKNIKDKAKIPFALDSEADPAKGEYSGACATIG 418  
 QY 292 TTGGVMEALRTAYELFTGTPRLSLSEVRGMDGICETNITWVPAPGSKFEELKRAA 351  
 DB 419 AAGGVMEALRSKADRAENAELEDIEYKQVRGNGIKFAVEI-----NNKKY----- 466  
 QY 352 AAEEAAAHGTPGLANDGAGFTSEDRGCTLRVAVANGSLGNAKKLITRKMAQGEAKYDF 411  
 DB 467 -----NNAVINGASNLFFPMKSGMINERKOHF 493  
 QY 412 VEIMACPACGVGGGGRSTDK-----ATTOKROALYMLDEKSTLRSHENPSIEELD 466  
 DB 494 IEVMACHGCGVNGGGPHVNPDKLEVDIKVNAASVLYMODEHLSKRKSHENTALVAKMQ 553  
 QY 467 TYLGEPLGKHAHELTHY 485  
 DB 554 NYFGKPGEGRAHEILHFY 572  
 RESULT 2  
 PHFL\_DESVH STANDARD; PRT; 421 AA.  
 ID PHFL\_DESVH  
 AC P07598;  
 DT 01-APR-1988 (Rel. 07, Created)  
 DT 01-APR-1988 (Rel. 07, Last sequence update)  
 DT 01-JUL-1993 (Rel. 26, Last annotation update)  
 DE Periplasmic (Fe) hydrogenase large subunit (EC 1.18.99.1) (Fe  
 hydrolase).  
 OS Desulfovibrio vulgaris (strain Hildenborough).  
 OC Bacteria; Proteobacteria; delta subdivision; Desulfovibrionaceae;

OC Desulfovibrio.  
 OX NCBI\_TaxID=882;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=85203856; PubMed=3886621;  
 RA Voodou G., Brenner S.;  
 RT "Nucleotide sequence of the gene encoding the hydrogenase from  
 Desulfovibrio vulgaris (Hildenborough).";  
 RL Eur. J. Biochem. 148:515-520(1985).  
 RN [2]  
 RP SEQUENCE OF 1-35.  
 RX MEDLINE=88106446; PubMed=3322275;  
 RA Prickl B.C., He S.H., Li C., Menon N., Choi E.S., Przydla A.E.,  
 RA Deravertan D.V., Peck H.D., Jr., Faugue G., Legall J., Teixeira M.,  
 RA Moura I., Moura J.J.G., Pail D., Huynh B.H.;  
 RT "Identification of three classes of hydrogenase in the genus,  
 Desulfovibrio.";  
 RL Biochem. Biophys. Res. Commun. 149:369-377(1987).  
 RN [3]  
 RP SEQUENCE OF 1-15 FROM N.A.  
 RA van den Berg W.A.M., Stokkermans J.P.W.G., van Dongen W.M.A.M.;  
 RL Submitted (Sep-1992) to the EMBL/Genbank/DBI databases.  
 CC -1- FUNCTION: MAY BE INVOLVED IN HYDROGEN UPTAKE FOR THE REDUCTION OF  
 CC SULFATE TO HYDROGEN SULFIDE IN AN ELECTRON TRANSPORT CHAIN.  
 CC CYTOCHROME C3 IS LIKELY TO BE THE PHYSIOLOGICAL ELECTRON CARRIER  
 CC FOR THE ENZYME.  
 CC -1- CATALYTIC ACTIVITY: 2 reduced ferredoxin + 2 H(+) -> 2 oxidized  
 CC ferredoxin + H(2).  
 CC -1- COFACTOR: BINDS THREE 4FE-4S CLUSTERS. TWO OF THESE CLUSTERS ARE  
 CC PROBABLY COORDINATED TO THE FERREDOXIN LIKE DOMAIN IN THE N-  
 CC TERMINAL OF THIS PROTEIN.  
 CC -1- SUBUNIT: COMPOSED OF TWO SUBUNITS.  
 CC -1- SUBCELLULAR LOCATION: Periplasmic.  
 CC -1- MISCELLANEOUS: [FE], [NIFE], AND [NIFES] HYDROGENASES APPEAR TO  
 CC REPRESENT THREE DISTINCT ENZYMES HAVING HYDROGENASE ACTIVITY.  
 CC -1- SIMILARITY: THE IRON-SULFUR CENTERS ARE SIMILAR TO THOSE OF  
 CC 'BACTERIAL-TYPE' 4FE-4S FERREDOXINS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: X02416; CAA26266.1; -  
 DR EMBL: Z15142; CAA78848.1; -  
 DR PIR: A24551; HDVFL.  
 DR HSSP: P29166; IFEH.  
 DR InterPro: IPR001450; 4FE4S\_Ferredoxin.  
 DR InterPro: IPR004108; Fe\_hyd\_lg\_C.  
 DR Pfam: PF00037; fer4; 2.  
 DR Pfam: PF02906; Fe\_hyd\_lg\_C; 1.  
 DR PROSITE: P500198; 4FE4S\_FERREDOXIN; 2.  
 KW Oxidoreductase; Electron transport; 4Fe-4S; Iron-sulfur;  
 KW Periplasmic.  
 FT METAL 35 35 IRON-SULFUR (4FE-4S) (BY SIMILARITY).  
 FT METAL 38 38 IRON-SULFUR (4FE-4S) (BY SIMILARITY).  
 FT METAL 41 41 IRON-SULFUR (4FE-4S) (BY SIMILARITY).  
 FT METAL 45 45 IRON-SULFUR (4FE-4S) (BY SIMILARITY).  
 FT METAL 66 66 IRON-SULFUR (4FE-4S) (BY SIMILARITY).  
 FT METAL 69 69 IRON-SULFUR (4FE-4S) (BY SIMILARITY).  
 FT METAL 72 72 IRON-SULFUR (4FE-4S) (BY SIMILARITY).  
 FT METAL 76 76 IRON-SULFUR (4FE-4S) (BY SIMILARITY).  
 FT METAL 77 77 IRON-SULFUR (4FE-4S) (BY SIMILARITY).  
 FT CONFLICT 35 35 C -> K (IN REF. 2).  
 SQ SEQUENCE 421 AA; 45951 MW; 8E70A0775B86AACD CRC64;  
 Query Match 24.6%; Score 633; DB 1; Length 421;  
 Best Local Similarity 36.6%; Pred. No. 6.9e-37;  
 Matches 149; Conservative 56; Mismatches 122; Indels 80; Gaps 9;  
 QY 54 CAAAPAAEAPLSHVOQALAEIAKPKDDPRKHYCVQVAPAVRAIAETLGLAPGATT 112

```

DB 72 CLTHCP--ENAIYEASWPEVEKIKDKVK--CIAMPAPAVRYALGAFGMPYGSVTT 127
OY 113 KOLAGLRLIGFDEVDTLFGADLTIMESGSELLHLEHLSHPSDPLMFETSCCG 172
DB 128 GKMLAALQGLGHAHCHEDTFTADVTIMESSEVERLTK-----KSDMPLQFETSCCG 181
OY 173 WTAMLEKSYDILIPYVSSCKSPOMLAAMVSKYLAERKCIAPKDMVMSIMPTCKOSBA 232
DB 182 WCKATETYPPELLPHSTCKSPIGNMGALAKYGAERMKYDKROYTVYSIMCIACKYGS 241
OY 233 DDDMFCVADPFL-----RDLHVITTVELGNIFERGINLAELPGEENDMPYGVSGAG 287
DB 242 LR-----PELKSSGMKDIDATLTTRRLAYMIKKAIGIDFAPGKRDSLMSGESTGGA 293
OY 288 VLEGTGGVMEALRFAVLELTGTPLRLSLSEVRNMDIKENITMVPAPSGFEELK 347
DB 294 TLEGTGGVMEALRFAVLELTGTPLRLSLSEVRNMDIKENITMVPAPSGFEELK 340
OY 348 HRAAARAEMAAGTPEGLAMDGAGFTSDGRGILRYAVANGLSNAKLLTKMOAGBA 407
DB 341 -----GCTDVKAVVHGAKRFQVCDVYAGKS 368
OY 408 KIDPEYIMACPRACVGGGQOP-----RSTDKATIQ-----KRAAL 443
DB 369 PYHFLIYMACPGCGVCGGQPYMPGVLEAMDRTRTLRYAGLKKRLAM 415

RESULT 3
PHFL_DESVO STANDARD; PRT; 421 AA.
ID PHFL_DESVO STANDARD; PRT; 421 AA.
AC P13629;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 01-MAR-1992 (Rel. 21, Last annotation update)
DE Periplasmic [Fe] hydrogenase large subunit (EC 1.18.99.1) (Fe
DE hydrogenlyase).
GN HYDA.
OS Desulfovibrio vulgaris (strain Oxamticus Monticello).
OC Bacteria; Proteobacteria; delta subdivision; Desulfovibrionaceae;
OC Desulfovibrio.
OX NCBI_TaxID=884;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89291738; PubMed=2661538;
RT Voordouw G., Strang J.D., Wilson F.R.:
RT "Organization of the genes encoding [Fe] hydrogenase in Desulfovibrio
RT vulgaris subsp. Oxamticus Monticello."
RT J. Bacteriol. 171:3881-3889(1989).
CC -1 FUNCTION: MAY BE INVOLVED IN HYDROGEN UPTAKE FOR THE REDUCTION OF
CC SULFATE TO HYDROGEN SULFIDE IN AN ELECTRON TRANSPORT CHAIN.
CC CYTOCHROME C3 IS LIKELY TO BE THE PHYSIOLOGICAL ELECTRON CARRIER
CC FOR THE ENZYME.
CC -1 CATALYTIC ACTIVITY: 2 reduced ferredoxin + 2 H(+) -> 2 oxidized
CC ferredoxin + H(2).
CC -1 COFACTOR: BINDS THREE 4FE-4S CLUSTERS. TWO OF THESE CLUSTERS ARE
CC PROBABLY COORDINATED TO THE FERREDOXIN LIKE DOMAIN IN THE N-
CC TERMINAL OF THIS PROTEIN.
CC -1 SUBUNIT: COMPOSED OF TWO SUBUNITS.
CC -1 SUBCELLULAR LOCATION: Periplasmic.
CC -1 MISCELLANEOUS: [Fe], [NIFE], AND [NIFES] HYDROGENASES APPEAR TO
CC REPRESENT THREE DISTINCT ENZYMES HAVING HYDROGENASE ACTIVITY.
CC -1 SIMILARITY: THE IRON-SULFUR CENTERS ARE SIMILAR TO THOSE OF
CC 'BACTERIAL-TYPE' 4FE-4S FERREDOXINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).

```

```

DR EMBL: M27212; AAA23373.1; .
DR PIR: A32866; HODVLY.
DR HSSP: P29166; 1FER.
DR InterPro: IPR001450; 4Fe4s_ferredoxin.
DR InterPro: IPR004108; Fe_hyd_19_C.
DR Pfam: PF00037; fer4; 2.
DR Pfam: PF02906; Fe_hyd_19_C; 1.
DR PRINTS: PR00353; 4FE4SFERDOXIN.
DR PROSITE: PS00198; 4FE4S_FERREDOXIN; 2.
KW Oxidoreductase; Electron transport; 4Fe-4S; Iron-sulfur;
KW Periplasmic.
KW METAL 35
FT METAL 35
FT METAL 38
FT METAL 41
FT METAL 45
FT METAL 66
FT METAL 69
FT METAL 72
FT METAL 76
SQ SEQUENCE 421 AA; 46278 MW; BE987ABC4DC7C965 CRC64;

Query Match 23.3%; Score 599.5; DB 1; Length 421;
Best Local Similarity 36.5%; Pred. No. 1,5e-34;
Matches 135; Conservative 49; Mismatches 117; Indels 69; Gaps 5;

OY 87 VCQVQVAVYVAIAETLGLAPGATPQKOLAEGRIGFDEVPDTLFGADLTIMESSELL 146
DB 102 VIAMPAPAVRYALGDAFGLPVGYVTGKMFSALEKGFCHDCHNETADVTIIEDETEY 161
OY 147 HRLTEHLAHPHSDEPLPMTSCCPGIMALEKSYDILIPYVSSCKSPOMLAAMVSKYL 206
DB 162 QRLTKKL-----DKPLPQGTSCCPGMHRYVESLYPELPHMSSCKSPIGMIGTLAKTYG 215
OY 207 AEKKGIAPDMVAVSIMPTCKROSEADR--WFCVADPRLRLDHYITVELGNIFKER 264
DB 216 ADMKRYDRAKVVYVSIPTCKRYEGRMRLM-----DSGKMDIDTITRRLAYMIKKA 270
OY 265 GINLAELPGEENDMPYGVSGAGVLFGTGGVMEALRFAVLELTGTPLRLSLSEVRGM 324
DB 271 KIDFTKLPQKRLTLGSESTGATLFGVGGVMEALRFAVAVTQVKKPESMPKVGRL 330
OY 325 DGIKETNITMVPAPSGKFEELKRAAARAEMAAGTPEGLAMDGAGFTSEDGRGTL 384
DB 331 QGVKEATVNV-----GGVVO 345
OY 385 RVAVANGLSNAKLLTRKMOGEAKYFVEIMACPRACVGGGQOP-----RSTDK 433
DB 346 KVAVVGARFRHDVCELVAKAPMHFIEMACPGCGVCGQPYMPGVLEADRRSTRM 405
OY 434 AIYQKRAAL 443
DB 406 YAGLKKRLAM 415

RESULT 4
NARL_YEAST STANDARD; PRT; 491 AA.
ID NARL_YEAST STANDARD; PRT; 491 AA.
AC P23503;
DT 01-NOV-1991 (Rel. 20, Created)
DT 01-NOV-1991 (Rel. 20, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Nuclear architecture related protein 1.
GN NARL OR YNL240C OR N114.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97051596; PubMed=8896273;
RA Pandolfi D., de Antoni A., Lanfranchi G., Valle G.;
RT "The DNA sequence of cosmid 14-5 from chromosome XIV reveals 21 open
RT reading frames including a novel gene encoding a globin-like

```

RT domain.\*;  
 RL Yeast 12:1071-1076(1996).  
 RN (12)  
 RP SEQUENCE OF 147-491 FROM N.A.  
 RX MEDLINE=91099671; PubMed=2269430;  
 RA Nogae I., Johnston M.;  
 RT "Isolation and characterization of the ZWF1 gene of Saccharomyces  
 cerevisiae, encoding glucose-6-phosphate dehydrogenase.\*"  
 RL Gene 96:161-169(1990).  
 CC -1- SIMILARITY: TO K.LACTIS LEFT AND S.POMBE SPOC1450.10C.  
 CC -1- CAUTION: REF.2 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO  
 CC FRAMESHIFTS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
 CC or send an email to license@isb-sib.ch).  
 CC -----  
 CC EMBL: 269381; CAA93358.1; -;  
 CC EMBL: 271516; CAA96145.1; -;  
 CC EMBL: M34709; AAA36618.1; ALT\_FRAME.  
 CC PIR: J00891; J00891.  
 CC SCD: S0005184; NARL.  
 CC InterPro: IPR004108; Fe\_hyd\_1g\_C.  
 CC Pfam: PF02906; Fe\_hyd\_1g\_C; 1.  
 SO SEQUENCE 491 AA; 54151 MW; FFE2A07A881B8C1 CRC64;

Query Match 9.9%; Score 255; DB 1; Length 491;  
 Best Local Similarity 23.1%; Pred. No. 1.6e-10;  
 Matches 114; Conservative 72; Mismatches 189; Indels 118; Gaps 18;

QY 52 VACAAAPAAEAPLASHVQ-----QALAEIARKKDDPTKRVQVAPARVAIAETLGL 105  
 DB 60 LACAGCTTSEEILLSSQSHVFLKMKWGLSQQD---KELVSVSPQCLSLADYYGL 115  
 QY 106 APCATTPKQAEGLRLGFEDEVFTLFGADLTME-----EGSEL-----LHRL 149  
 DB 116 -----TLEADLCMNEFOKHFQCKYVGMTEGHLISKT 151  
 QY 150 TEHLEAHPHSDEPL-----PMFTSCCPGMIALEKSTPDLIPYSSCKSQMMLAAVKS 204  
 DB 152 VERITIAHKKOKENTGADRKPLLSAVCPGFLYTEKRPOLVPMILNVKSPQOITGSLIRA 211  
 QY 205 YLAEEKGIAPKDMVMSIMPTCKROSEADRMFCVADPTLROLDHYITVETGNIKFKER 264  
 DB 212 TPESLAIAHESFYHLSIMPCFQKLEASRPESLDG-----IDCVITTRVETMLQEL 264  
 QY 265 GINLAEL-----PEGEMDNPMSGVSGAGVLEGTGGVMEALRTAV ELFTGT- 311  
 DB 265 NLDKFSFLTEDTSLYGLRSPG--WDPVHMASNLG---GTCGGY-----AVQYTAVQ 313  
 QY 312 ---PLPRLSLSEVRGMDIGETNITWVPAPGSKFEELKLR---MAARAAMAHGTGPL 365  
 DB 314 RLHPGSMIVLEGRNSDIY-----EYRLLDHDDRTIAAASLSGFRNIQNL 368  
 QY 366 AMDGAGFTSEDRGIGITLRAVAVANGLNKKLITKQAGEAKY--DFVEIMACPGAGVG 423  
 DB 359 VRKLTSGSGSEKRRNITALKRKRTGPKANSREMAAATAADPHISDYIEVNAACPGACMN 418  
 QY 424 GGGCPRSTDAITQKROALYNIIDEKSTLRSHENSIRELYDYTLGEPLGHAHELLHT 463  
 DB 419 GGGGLNEQNSLRKQVY-----QTLNKRH--GEELANVDDPLTLGPKLEEAARLSTL 469  
 QY 484 HYVAGVEEKDEK 496  
 DB 470 EYVFAPYKQAVK 482

RESULT 5  
 LEFT\_KLUOLA

ID LEFT1\_KLUOLA STANDARD; PRT; 469 AA.  
 AC P53998;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 01-OCT-1996 (Rel. 34, Last annotation update)  
 DE LEFT1 protein.  
 GN LEFT1.  
 OS Kluyveromyces fragilis (Yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; Kluyveromyces.  
 OX NCBI\_TaxID=28985;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CBS 2359/152;  
 RA Mesolowski-Louyel M., Tanguy-Rougeau C., Fukuhara H.;  
 RL Submitted (FEB-1993) to the EMBL/GenBank/DBJ databases.  
 CC -1- SIMILARITY: TO YEAST YNL240C.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
 CC or send an email to license@isb-sib.ch).  
 CC -----  
 CC EMBL: X70373; CAA49833.1; -;  
 CC InterPro: IPR004108; Fe\_hyd\_1g\_C.  
 CC Pfam: PF02906; Fe\_hyd\_1g\_C; 1.  
 SO SEQUENCE 469 AA; 52072 MW; 555DAAE7DEB9B9B7 CRC64;

Query Match 9.0%; Score 232; DB 1; Length 469;  
 Best Local Similarity 23.9%; Pred. No. 6.1e-09;  
 Matches 110; Conservative 62; Mismatches 167; Indels 122; Gaps 16;

QY 52 VACAAAPAAEAPLASHVQALAEIARKKDDPTKRVQVAPARVAIAETLGLAPGA 109  
 DB 57 LACAGCTTSEEILLSSQSHVFLKMKWGLSQQD---KELVSVSPQCLSLADYYSMC--- 113  
 QY 110 TTPKQAEGLR-----RLGFEDEVFTLFGADLTMEEGSELHRLTEHLEAHPHSDE 161  
 DB 114 ---LADDRCONPFAKTFNKNKYVGTGFGSISIRNATLTKRVPEN----- 159  
 QY 162 PLDMFTSCCPGMIALEKSTPDLIPYSSCKSQMMLAAVKSYLEKKGIAPKDMVMS 221  
 DB 160 EGPFLSCVCPGFVLYEKTRELPIMLDVKSPOQITGSLIRK-----ADPTCYHLS 211  
 QY 222 IMPTCKROSEADRMFCVADPTLROLDHYITVETGNIKFKEGINL-----AEL 271  
 DB 212 IMPCFQKLEASR-----EBCREYDCVITTRKQFVAMLGDSLIDKSTVTEYDSSKEL 264  
 QY 272 PEGEMDNPMSGVSGAGVLEGTGGVMEALR-----PAYELFTG 310  
 DB 265 CPFGMDYKHLMLNMG---SSSGYAVQYLLSSQSNPESDITTEGKNSDVEYRLVSK 321  
 QY 311 TPLPRLSLSEVRGMDI-----KETNITWVPAPGSKFEELKIRAAARAAMAHGTGPL 365  
 DB 322 SKGVVIASSSEVYFRNIQNLVRKLSQASVKKKGIK---VKRRGOSVLR----- 367  
 QY 366 AMDGAGFTSEDRGIGITLRAVAVANGLNKKLITKQAGEAYVDFEIMACPGAGVG 425  
 DB 368 ---SGETSE-----KTKVLT---ADPATDVEVNAACPSGCGINGG 402  
 QY 426 GGPSTDAITQKROALYNIIDEKSTLRSHENSIRELYD 466  
 DB 403 GL-LNEEKNANRRKQLA-QDLSLAYTKVHVSVINPDYHAYD 441

RESULT 6  
 PHFS\_DESVH  
 ID PHFS\_DESVH STANDARD; PRT; 123 AA.  
 AC P07603;  
 DT 01-APR-1988 (Rel. 07, Created)

DT 01-APR-1988 (Rel. 07, Last sequence update)  
 DT 01-MAY-1991 (Rel. 18, Last annotation update)  
 DE Periplasmic [Fe] hydrogenase small subunit precursor (EC 1.18.99.1)  
 GN (Fe hydrogenylase small chain).  
 OS Desulfovibrio vulgaris (strain Hildenborough).  
 CC Bacteria; Proteobacteria; delta subdivision; Desulfovibrionaceae;  
 CC Desulfovibrio.  
 CC NCB1\_TaxID=882;  
 RX MEDLINE=85203856; PubMed=388621;  
 RA Voordouw G., Brenner S.;  
 RT "Nucleotide sequence of the gene encoding the hydrogenase from  
 Desulfovibrio vulgaris (Hildenborough).";  
 RL Eur. J. Biochem. 148:515-520(1985).  
 RP SEQUENCE FROM N.A.  
 RP MEDLINE=86277938; PubMed=3525521;  
 RA Prickril B.C., Czechowski M.H., Przybyla A.E., Peck H.D. Jr.,  
 RA Le Gall J.;  
 RT "Putative signal peptide on the small subunit of the periplasmic  
 hydrogenase from Desulfovibrio vulgaris.";  
 RL J. Bacteriol. 167:722-725(1986).  
 CC -1- FUNCTION: MAY BE INVOLVED IN HYDROGEN UPTAKE FOR THE REDUCTION OF  
 CYTOCHROME C3 IS LIKELY TO BE THE PHYSIOLOGICAL ELECTRON CARRIER  
 FOR THE ENZYME.  
 CC -1- CATALYTIC ACTIVITY: 2 reduced ferredoxin + 2 H(+) = 2 oxidized  
 ferredoxin + H(2).  
 CC -1- SUBUNIT: HETERODIMER OF A LARGE AND A SMALL SUBUNIT.  
 CC -1- SUBCELLULAR LOCATION: Periplasmic.  
 CC -1- MISCELLANEOUS: [FE], [NIFE], AND [NIFESF] HYDROGENASES APPEAR TO  
 REPRESENT THREE DISTINCT ENZYMES HAVING HYDROGENASE ACTIVITY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: X02416; CAA26267.1; -  
 DR PIR: B24551; HODVSV.  
 DR InterPro: IPR003149; Fe\_hyd\_SSU.  
 DR Pfam: PF02256; Fe\_hyd\_SSU; 1.  
 KW Oxidoreductase; Periplasmic; Iron-sulfur; Signal.  
 FT SIGNAL 1 34  
 FT CHAIN 35 123 PERIPLASMIC [FE] HYDROGENASE SMALL  
 FT SUBUNIT.  
 SQ SEQUENCE 123 AA; 13624 MW; 2F4F7A4304ECC47B CRC64;  
 Query Match 5.4%; Score 138.5; DB 1; Length 123;  
 Best Local Similarity 38.7%; Pred. NO. 0.0037;  
 Matches 29; Conservative 14; Mismatches 21; Indels 11; Gaps 2;  
 QY 429 RSTKAITOKRQA-----ALYNLDEKSTLRSHNPSTRELYDYLGSEPLCKRAHE 479  
 DB 27 RRTGAAVAARQIKDYMLDRINGVYGAADAKFPVRSQDNTQVAKLYSKYLEPLCKSHD 86  
 QY 480 LLHTHYV--AGGVEE 492  
 DB 87 LLHTHMFDSKSGVE 101  
 RESULT 7  
 PHFS\_DESVO STANDARD: PRT: 124 AA.  
 AC P13628;  
 DT 01-JAN-1990 (Rel. 13, Created)  
 DT 01-JAN-1990 (Rel. 13, Last sequence update)  
 DT 01-MAY-1991 (Rel. 18, Last annotation update)

DE Periplasmic [Fe] hydrogenase small subunit precursor (EC 1.18.99.1)  
 DE (Fe hydrogenylase small chain).  
 GN HYDB.  
 OS Desulfovibrio vulgaris (strain Oxamimus Monticello).  
 CC Bacteria; Proteobacteria; delta subdivision; Desulfovibrionaceae;  
 CC Desulfovibrio.  
 CC NCB1\_TaxID=884;  
 RX MEDLINE=89291738; PubMed=2661538;  
 RA Voordouw G., Strang J.D., Wilson F.R.;  
 RT "Organization of the genes encoding [Fe] hydrogenase in Desulfovibrio  
 vulgaris subsp. Oxamimus Monticello.";  
 RL J. Bacteriol. 171:3881-3889(1989).  
 CC -1- FUNCTION: MAY BE INVOLVED IN HYDROGEN UPTAKE FOR THE REDUCTION OF  
 SULFATE TO HYDROGEN SULFIDE IN AN ELECTRON TRANSPORT CHAIN.  
 CC CYTOCHROME C3 IS LIKELY TO BE THE PHYSIOLOGICAL ELECTRON CARRIER  
 FOR THE ENZYME.  
 CC -1- CATALYTIC ACTIVITY: 2 reduced ferredoxin + 2 H(+) = 2 oxidized  
 ferredoxin + H(2).  
 CC -1- SUBUNIT: HETERODIMER OF A LARGE AND A SMALL SUBUNIT.  
 CC -1- SUBCELLULAR LOCATION: Periplasmic.  
 CC -1- MISCELLANEOUS: [FE], [NIFE], AND [NIFESF] HYDROGENASES APPEAR TO  
 REPRESENT THREE DISTINCT ENZYMES HAVING HYDROGENASE ACTIVITY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: M27212; AAA23374.1; -  
 DR PIR: B32886; HODVSV.  
 DR InterPro: IPR003149; Fe\_hyd\_SSU.  
 DR Pfam: PF02256; Fe\_hyd\_SSU; 1.  
 KW Oxidoreductase; Periplasmic; Iron-sulfur; Signal.  
 FT SIGNAL 1 34  
 FT CHAIN 35 124 PERIPLASMIC [FE] HYDROGENASE SMALL  
 FT SUBUNIT.  
 SQ SEQUENCE 124 AA; 13962 MW; 296AF34FD32B21B5 CRC64;  
 Query Match 5.0%; Score 130; DB 1; Length 124;  
 Best Local Similarity 36.2%; Pred. NO. 0.014; 26; Indels 12; Gaps 2;  
 Matches 29; Conservative 13; Mismatches 26;  
 QY 419 AGCVGGG---OPRSTDKAITOKRQA-----ALYNLDEKSTLRSHNPSTRELYD 466  
 DB 14 AACVVTGGLISIRMTGRANAARAKQDKYMDRINQVYGAADAKFPVRSQDNTQVAKLYA 73  
 QY 467 TYLGEPLGKHAHELHTHYV 486  
 DB 74 DELERKMSHKAQLHTHWV 93  
 RESULT 8  
 HYPF\_AZOVI STANDARD: PRT: 776 AA.  
 AC P40596;  
 DT 01-FEB-1995 (Rel. 31, Created)  
 DT 15-JUL-1998 (Rel. 36, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE Hydrogenase maturation protein hypf.  
 GN HYPF.  
 OS Azotobacter vinelandii.  
 CC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;  
 CC Azotobacter.  
 CC NCB1\_TaxID=354;  
 RX MEDLINE=92305060; PubMed=1610901;

RA Chen J.C., Mortenson L.E.;  
 RT Identification of six open reading frames from a region of the  
 RT Azotobacter vinelandii genome likely involved in dihydrogen  
 RT metabolism";  
 RL Biochim. Biophys. Acta 1131:199-202(1992).  
 CC  
 CC -1- FUNCTION: INVOLVED IN THE HYDROGENASE MATURATION PROCESS (BY  
 CC SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO THE HYFP FAMILY.  
 CC -1- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO THE  
 CC ACYLPHOSPHATASE FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; X63650; CAA45185.1; ALT\_INIT.  
 CC EMBL; L23970; AAA19510.1; ALT\_INIT.  
 CC Interpro: IPR001792; Acylphosphatase.  
 CC Interpro: IPR004421; Hyfp.  
 CC Interpro: IPR000666; Sua5\_ycl10\_yrdc.  
 CC Pfam: PF00708; Acylphosphatase; 1.  
 CC Pfam: PF01300; Sua5\_ycl10\_yrdc; 1.  
 CC Prodom: PD001884; Acylphosphatase; 1.  
 CC Prodom: PD002209; Sua5\_ycl10\_yrdc; 1.  
 CC Trigram: TIGR00143; hyfp; 1.  
 CC PROSITE; PS00150; ACYLPHOSPHATASE\_1; 1.  
 CC ZINC-finger.  
 CC  
 CC DOMAIN 1 93 ACYLPHOSPHATASE-LIKE.  
 CC ZN\_FING 113 138 C4-TYPE (POTENTIAL).  
 CC FT ZN\_FING 163 188 C4-TYPE (POTENTIAL).  
 CC SQ SEQUENCE 776 AA; 82732 MW; E997337C0A15600C CRC64;  
 CC  
 CC Query Match 4.38; Score 112; DB 1; Length 776;  
 CC Best Local Similarity 21.8%; Pred. No. 2.5;  
 CC Matches 130; Conservative 53; Mismatches 220; Indels 192; Gaps 28;  
 CC  
 CC 29 PLASIVRVATLLEAPRLGN--VACAAAPAAEAPLS-----HV-----QQAIA 73  
 CC 182 PLICPACGPRLLCEDSGRLSDPIERLALALAREGIELLRGVGFFHLCDARNAGAVA 241  
 CC 74 ELAKPKDDPTKRVCAQVAPAVNAIAE--TLGIA--PGATPKQIAEGRLRGFEVFD 129  
 CC 242 ELRRRRRRPAKPPALMAANASIALQALVELNENGLAELSGRAAVVLLR--KRAAADLLA 259  
 CC 130 TLFGADLTTEESESELHRLTEHLLEAPHSDEPLPMFTSCPGWIAMLEKSYDILPYVS 189  
 CC 300 AGVAPDLAWL--GAMLPSPHLHLLEFHEAAGR-----AGTQMAA----- 337  
 CC 190 SCSPQWMLAAMKSYAEKKGIAPKDMWVSIPTCRKSEADR-----DMFCVDADP 243  
 CC 338 ---APQDLVMTSANISG---AP-----PTGNAREBERLAGIDLHLHRET 361  
 CC 244 TLRLDHLVITVELGNIFKEKGINLAELPGEWDMPMVGSGAVLEGGTGVMEALRT 303  
 CC 382 LKNCDDSVSALBRAPLVIRSGRLAPL-----EPLARG--GPSIL--ALGG-----QLKN 429  
 CC 304 AYELEGTPLPLSLSEVRG-----MDGIKETNTMVPAP----- 338  
 CC 430 AVCLTRG--DRAMLSPHNDLDDADTCLALERTVARLGLGIRPERVACDLHPDFAS 486  
 CC 339 -----GSKFEELKRAARAARAAAHGTRCP---LAWG--GAGFTSEDS--RGSI 382  
 CC 487 RFAADYAEHRGLPLPHQHHAHIAATMAEHGLAEPVGLALDGFGLG---ADGRURGE 543  
 CC 383 TLRYA-----VANGIAGNAKKLITKMGAGEAKYDF 411  
 CC 544 LLRYAADGCAMLGELNPLPLPGDDQASREPWRMAAGLHLNGGELATRFAPAEPGAAT 603  
 CC 412 VEIWA-----CP-----AGCVGGGGGPRSDKAITO-----KKAQALYNL 446

DB 604 ARMLERGFNCPPSSSAGRLFDAAAGLIGGRQREAEALRLSEKRVGLPERVPALMRI 663  
 QY 447 DEKSTLRSHENDSIKELYDTYLGEPIG-----KAHELLTHRYVAG 489  
 DB 664 DDDNRLLPLPLARLADMDDTRAGAELEFGVLEALATMAARAAMETGILRYALG 718  
 CC  
 CC RESULT 9  
 CC ICAS RABIT  
 CC ID ICAS RABIT STANDARD; PRT; 912 AA.  
 CC AC Q28730;  
 CC DT 16-OCT-2001 (Rel. 40, Created)  
 CC DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 CC DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 CC DE Intercellular adhesion molecule-5 precursor (ICAM-5) (telencephalin).  
 CC GN ICAM-5 OR TCN.  
 CC OS Oryctolagus cuniculus (Rabbit).  
 CC OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 CC OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.  
 CC OX NCBI\_Taxid=9986;  
 CC RN [1]  
 CC RP SEQUENCE FROM N.A.  
 CC RC STRAIN=Japanese white; TISSUE=Brain;  
 CC RX MEDLINE=94206354; PubMed=7794412;  
 CC RA Yoshitake Y., Oka S., Nemoto Y., Watanabe Y., Nagata S.,  
 CC RA Kagamiyama H., Mori K.;  
 CC RT "An ICAM-related neuronal glycoprotein, telencephalin, with brain  
 CC RT segment-specific expression.";  
 CC RL Neuron 12:541-553(1994).  
 CC CC  
 CC -1- FUNCTION: ADHESION MOLECULE THAT BINDS TO LEUKOCYTE ADHESION  
 CC LFA-1 PROTEIN (INTEGRIN ALPHA-L/BETA-2).  
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.  
 CC -1- TISSUE SPECIFICITY: Expressed on neurons in the most rostral  
 CC segment of the mammalian brain, the telencephalon.  
 CC -1- SIMILARITY: BELONGS TO THE IMONOGLOBULIN SUPERFAMILY. ICAM  
 CC SUPERFAMILY.  
 CC -1- SIMILARITY: CONTAINS 9 IMONOGLOBULIN-LIKE C2-TYPE DOMAINS.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; L31399; AAA18478.1; -  
 CC HSP; P05362; IIC1.  
 CC Interpro: IPR003988; ICAM.  
 CC Interpro: IPR003987; ICAM\_VCAM-1.  
 CC Interpro: IPR003006; IG\_MHC.  
 CC Interpro: IPR003598; IG\_C2.  
 CC Interpro: IPR003600; Ig\_Like.  
 CC Pfam: PF00047; Ig; 6.  
 CC PRINTS; PRO1473; ICAM.  
 CC DR PRINTS; PRO1472; ICAMVCAI.  
 CC SMART; SM00410; IG\_Like; 3.  
 CC SMART; SM00408; IGC2; 1.  
 CC Immunoglobulin domain; Cell adhesion; glycoprotein; Transmembrane;  
 CC Repeat; Signal.  
 CC KW Repeat; Signal.  
 CC FT SIGNAL 1 29 POTENTIAL.  
 CC FT CHAIN 30 912 INTERCELLULAR ADHESION MOLECULE-5.  
 CC FT DOMAIN 30 826 EXTRACELLULAR (POTENTIAL).  
 CC FT TRANSSEM 827 847 POTENTIAL.  
 CC FT DOMAIN 848 912 CYTOPLASMIC (POTENTIAL).  
 CC FT DOMAIN 47 107 IG-LIKE C2-TYPE DOMAIN 1.  
 CC FT DOMAIN 132 202 IG-LIKE C2-TYPE DOMAIN 2.  
 CC FT DOMAIN 239 304 IG-LIKE C2-TYPE DOMAIN 3.  
 CC FT DOMAIN 332 385 IG-LIKE C2-TYPE DOMAIN 4.  
 CC FT DOMAIN 403 472 IG-LIKE C2-TYPE DOMAIN 5.  
 CC FT DOMAIN 486 552 IG-LIKE C2-TYPE DOMAIN 6.  
 CC FT DOMAIN 566 645 IG-LIKE C2-TYPE DOMAIN 7.

FT	DOMAIN	659	724	IG-LIKE C2-TYPE DOMAIN 8.
FT	DOMAIN	754	813	IG-LIKE C2-TYPE DOMAIN 9.
FT	DISULFID	54	101	BY SIMILARITY.
FT	DISULFID	139	195	BY SIMILARITY.
FT	DISULFID	246	297	BY SIMILARITY.
FT	DISULFID	339	378	BY SIMILARITY.
FT	DISULFID	430	465	BY SIMILARITY.
FT	DISULFID	546	546	BY SIMILARITY.
FT	DISULFID	573	638	BY SIMILARITY.
FT	DISULFID	666	717	BY SIMILARITY.
FT	DISULFID	761	806	BY SIMILARITY.
FT	CARBOHYD	53	53	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	134	134	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	192	192	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	211	211	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	311	311	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	366	366	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	392	392	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	576	576	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	639	639	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	756	756	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	787	787	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	788	788	N-LINKED (GLCNAC. . .) (POTENTIAL).
SO	SEQUENCE	912 AA;	95682 MM;	4DC032853DA0EAE9 CRC64;
Query Match				
Best Local Similarity 4.3%; Score 111; DB 1; Length 912;				
Matches 114; Conservative 56; Mismatches 210; Indels 166; Gaps 26;				
OY	18	SCRANOVAPRAPLAASVVALATLE----	APARRLGNVCAAAAPAE-----	A 63
DB	138	SCRVPAGBGRSLTLTLRGAOELIRSFAGEPARAGAVLAVTLARREDHCANFSCRA		197
OY	64	PLSHVQQA-----	ELAKPKDPTKKNHCYVAVRAIAETL-----	GL 105
DB	198	ELDLRPGGLALEFNSSAPROLMTYALPLDSPR-----	LAPVLEVDSSQSLVSCITDGI	251
OY	106	ARGATTPKQALAEIRLGFDEVDTLFG-----	ADLTIMEEG-SELLHRLT-----	BHLEA 155
DB	252	FRASEGVHLAGDKRLN-PEV--	ILEGAIYATATATAEBSIKOLVCATVIGGERRES	308
OY	156	HPH---SDEPLPMFTSCCPGWTAMLEKSYPDLPYVSSCKSPQMLAAMVKSYLEKKG1		212
DB	309	RENVTYSSPAPLLT-----	LSEPSAPEKGLVTCTAGARALVTL-----	EGV 352
OY	213	AKDAMWMSIMPTKRSQSRADRDPCVADPTIRQIDHVT---	TYELGNIFERGINLA	269
DB	353	PAAPGQPAQLQFNNSDDGRSFC--	DATLELDGETLSKNGSAELRYLAPR-LDDA	408
OY	270	ELPEGEWDPNG-----	VSGAGVLFPTGCVMEALRTAYELF	308
DB	409	DCPR-SWMPPEGPQTLRCBARGNPTPAVHCARSDGAVATLCLLPRTALAGTY----		463
OY	309	TGTPPLRLSLSEVRGMDGIKETNITVPAP-----	GSKEELLKHAARAAEAHAAGTGC	363
DB	464	-----RCYAAANYQG-EAVKDVTLTVEYAPALDVSVCPEKVTVEGTEASLSCVANGVP		516
OY	364	PLAMDGAGFTSDGNGI-----	TLRAVANGLGNAKLLITKMGAGEKTYFVEI	414
DB	517	PSV-SCVRFROADVLEGLLVARAHAGTYRCEAIARALAKNAVATVEGSPS----	FEE	570
OY	415	MACPAG-----	CVGGGG-----	OPRSTDKAITOK 438
DB	571	RSPFSWMTWVEGSEQLFSCGEVECKPPSPVOCVSGEASBGLLPLAPLNPSPSPSVPRD		630
OY	439	ROAALY 444		
DB	631	LAPGIT 636		

RESULT 10  
CLPB\_ECOLI  
ID CLPB\_ECOLI STANDARD; PRT; 857 AA.

AC	P03815;
DT	21-JUL-1986 (Rel. 01, Created)
DT	01-NOV-1997 (Rel. 35, Last sequence update)
DT	15-JUN-2002 (Rel. 41, Last annotation update)
DE	ClpB protein (Heat shock protein F84.1).
GN	ClpB OR HTPW OR B2592 OR Z3886 OR EC53455.
OS	Escherichia coli, and
OS	Escherichia coli O157:H7.
OC	Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC	Psychrieta.
OX	NCBI_Taxid=562, 83334;
RN	[1]
RP	SEQUENCE FROM N.A.
RX	MEDLINE-90239044; PubMed-2185473;
RA	Gottesman S., Squires C., Pichersky E., Carlington M., Hobbs M.,
RA	Matick J.S., Dallymple B., Kuramitsu H., Shiroza T., Foster T.,
RA	Clark W.P., Ross B., Squires C.L., Maurizi M.R.;
RT	"Conservation of the regulatory subunit for the Clp ATP-dependent
RT	protease in prokaryotes and eukaryotes.";
RL	Proc. Natl. Acad. Sci. U.S.A. 87:3513-3517(1990).
RP	[2]
RP	SEQUENCE FROM N.A.
RC	STRAIN-K12 / MG1655;
RX	MEDLINE-97426617; PubMed-9278503;
RA	Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA	Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA	Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA	Mau B., Shao Y.;
RT	"The complete genome sequence of Escherichia coli K-12.";
RT	Science 277:1453-1474(1997).
RL	[3]
RP	SEQUENCE FROM N.A.
RC	STRAIN-K12;
RX	MEDLINE-97349890; PubMed-9205837;
RA	Yamamoto Y., Alba H., Baba T., Hayashi K., Inada T., Isono K.,
RA	Itoh T., Kimura S., Kitagawa M., Makino K., Miki T., Mitsuhashi N.,
RA	Mizobuchi K., Mori H., Nakade S., Nakamura Y., Nishimoto H.,
RA	Oshima T., Oyama S., Saito N., Sempel G., Satoh Y., Sivasubraman S.,
RA	Tagamori H., Takahashi H., Takeda J., Takemoto K., Uehara K., Wada C.,
RA	Yamagata S., Horinouchi T.;
RT	"Construction of a contiguous 874-kb sequence of the Escherichia coli
RT	- K12 genome corresponding to 50.0-68.8 min on the linkage map and
RT	analysis of its sequence features.";
RL	DNA Res. 4:91-113(1997).
RP	[4]
RP	SEQUENCE FROM N.A.
RC	STRAIN-O157:H7 / EDL933 / ATCC 700927;
RX	MEDLINE-21074935; PubMed-11206551;
RA	Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA	Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA	Postel G., Hackett J., Klink S., Boulton A., Shao Y., Miller L.,
RA	Grobleck E.J., Davis N.W., Lim A., Dimantata E.T., Potamousis K.,
RA	Apodaca A., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RT	Weich R.A., Blattner F.R.;
RT	"Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";
RT	Nature 409:529-533(2001).
RL	[5]
RP	SEQUENCE FROM N.A.
RC	STRAIN-O157:H7 / RIMD 0509952;
RX	MEDLINE-21156231; PubMed-11258796;
RA	Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA	Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tohe T.,
RA	Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
RA	Kihara S., Shiba T., Hattori M., Shinagawa H.;
RT	"Complete genome sequence of enterohaemorrhagic Escherichia coli
RT	O157:H7 and genomic comparison with a laboratory strain K-12.";
RL	DNA Res. 8:11-22(2001).
RP	[6]
RP	SEQUENCE OF 1-593 FROM N.A.
RC	STRAIN-K12;
RX	MEDLINE-91294165; PubMed-1906060;
RA	Kitagawa M., Wada C., Yoshioaka S., Yura T.;
RT	"Expression of ClpB, an analog of the ATP-dependent protease



RA Borselli D., Bowman S., Bruckner M., Carpenter J., Cherry J.M.,  
RA Chung E., Churcher C.M., Coster F., Davis K., Davis R.W.,  
RA Dietrich F.S., Delius H., DiPaolo T., Dubois E., Duesterhoeft A.,  
RA Duncan M., Floeth M., Fortin N., Friesen J.D., Fritz C., Goffeau A.,  
RA Hall J., Hebling U., Heumann K., Hilbert H., Hillier L.,  
RA Hunnicke-Smith S., Hyman R., Johnston M., Kallman S., Kleine K.,  
RA Komp C., Kurdi O., Lashkari D., Lew H., Lin A., Lin D., Louis E.J.,  
RA Marathe R., Messenguy F., Mewes H.-W., Mitipati S., Mostl D.,  
RA Mueller-Aber S., Narmath A., Nentwich U., Oefner P., Pearson D.,  
RA Petel F.X., Pohl T.M., Purnelle D., Schaefer M., Schaefer M.,  
RA Schreiers B., Schramm S., Schroeder M., Sdicu A.M., Tettein H.,  
RA Urrestearzu L.A., Ushinsky S., Verendeeels F., Vissers S., Voss H.,  
RA Walsh S.V., Wambut R., Wang Y., Wedler E., Wedler H., Winnett E.,  
RA Zhong W.Y., Zollner A., Vo D.H., Han J.,  
RA "The nucleotide sequence of Saccharomyces cerevisiae chromosome XVI,"  
RL Nature 387:103-105(1997).  
RN [4]  
RP MYOGENESIS OF GLY-1250.  
RC STRAIN-S228C;  
RX MEDLINE-94316198; PubMed-8041367;  
RA Inokoshi J., Tomoda H., Hashimoto H., Watanabe A., Takeshima H.,  
RA Omura S.;  
RT "Cerulenin-resistant mutants of Saccharomyces cerevisiae with an  
RT altered fatty acid synthase gene.";  
RL Mol. Gen. Genet. 244:90-96(1994).  
CC -1- FUNCTION: FATTY ACID SYNTHETASE CATALYZES THE FORMATION OF  
CC LONG-CHAIN FATTY ACIDS FROM ACETYL-COA, MALONYL-COA AND NADPH.  
CC THE ALPHA SUBUNIT CONTAINS DOMAINS FOR: ACYL CARRIER PROTEIN,  
CC 3-OXOACYL-[ACYL-CARRIER PROTEIN] REDUCTASE, AND 3-OXOACYL-[ACYL-  
CC CARRIER-PROTEIN] SYNTHASE. THIS SUBUNIT COORDINATES THE BINDING  
CC OF THE SIX BETA SUBUNITS TO THE ENZYME COMPLEX.  
CC -1- CATALYTIC ACTIVITY: Acetyl-CoA + N malonyl-CoA + 2N NADPH = 2  
CC long-chain fatty acid + (N+1) CoA + N CO(2) + 2N NADP(+).  
CC -1- CATALYTIC ACTIVITY: Acyl-[N+1]-carrier protein) + malonyl-[acyl]-  
CC carrier protein) = 3-oxoacyl-[acyl-carrier protein] + CO(2) +  
CC [acyl-carrier protein].  
CC -1- CATALYTIC ACTIVITY: (3n)-3-hydroxyacyl-[acyl-carrier protein] +  
CC NADP(+) = 3-oxoacyl-[acyl-carrier protein] + NADPH.  
CC -1- SUBUNIT: [Alpha(6)beta(6)] hexamers of two multifunctional  
CC subunits (alpha and beta).  
CC -1- SIMILARITY: TO THE FATTY ACID SYNTHETASE, SUBUNIT ALPHA FROM  
CC OTHER FUNGI.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its use  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; J03936; AAA34601.1; -;  
DR EMBL; X76890; CAA54218.1; -;  
DR EMBL; X94561; CAA64256.1; -;  
DR EMBL; Z73586; CAA97947.1; -;  
DR EMBL; Z73587; CAA97948.1; -;  
DR PIR; A31107; A31107.  
DR SGD; S0006152; FAS2.  
DR InterPro; IPR002582; ACPs.  
DR InterPro; IPR000794; Ketoacyl-synt.  
DR InterPro; IPR004568; Pantethn.tn.  
DR InterPro; IPR003880; Pantethn.tn.  
DR Pfam; PF00109; ketoacyl-synt.1.  
DR Pfam; PF01648; ACPs.1.  
DR Pfam; PF02801; ketoacyl-synt.C.1.  
DR ProDom; PD004282; ACPs.1.  
DR TIGRFAMs; TIGR00556; pantethn.tn.1.  
DR PROSITE; PS00012; PROSPOANTHETHEINE; 1.  
DR PROSITE; PS00606; B-KETOACYL-SYNTHASE; 1.  
KW Fatty acid biosynthesis; Multifunctional enzyme; Oxidoreductase;  
KW Transferrase; NADp; Phosphopantetheine.  
FT DOMAIN 1 ? ACYL CARRIER (ACP).  
FT 675 874 BETA-KETOACYL REDUCTASE.

FT	DOMAIN	1149	1363	BETA-KETOACYL SYNTHASE. PHOSHOANANTHETINE (BY SIMILARITY).
FT	ACT_SITE	1305	1305	BETA-KETOACYL SYNTHASE (BY SIMILARITY).
FT	MUTAGEN	1250	1250	G->S; CERULEIN-RESISTANCE.
FT	CONFLICT	310	310	G->G; GTGTGG (IN REF. 1).
FT	CONFLICT	594	594	T->I (IN REF. 1).
FT	CONFLICT	941	1019	AKIRFELVETSEKRAVSLTEPLAEHKVYVNSAAYNOVE TOPRANIOIDPELPKPYQVOKIOIAFAELBGLDLERYI -> CLNCKVSKMLKLTKEKOPPSKILMSIRLSMAIALAHMLKLS KENOELTFNPMTSQNRNHTNRRLKRLPLSLRWCMWIKELP (IN REF. 1).
FT	CONFLICT	1036	1041	RHEMCA->KMGNGS (IN REF. 1).
FT	CONFLICT	1408	1408	A->S (IN REF. 1).
FT	CONFLICT	1671	1671	N->T (IN REF. 1).
FT	SEQUENCE	1887 AA;	206946 MW;	08B872734CF3ABEA CRC64;
Query Match		4.2%;	Score 108.5;	DB 1; Length 1887;
Best Local Similarity		23.2%;	Pred. No. 13;	
Matches		67;	Conservative 33;	Mismatches 116; Indels 73; Gaps 12
QY	21	ARQVAPRAPIAASTRYVALATLEAPARRLGNVCAAPAAEAPLSHQOALAEAKPRD	80	
DB	104	AKKEAP-APTPAASAPAPAAAAAPAV-----AAAAPAA-----AAAEIA---D	142	
QY	81	DPTKHNQVOVAPARVAIAETLIGLAPGATTPKOLAEGRLRGPDEVFPTLGADUTIME	140	
DB	143	EPVAKSLLLHLVLA-----HKLKSLDSTIPMSKTKITDVGKSTVO- 183		
QY	141	EGSELLHRLTEHLEAHPHSDPELPM-----FTSCCPGIMALEKSYPLIIPYVSSCKSP	194	
DB	184	--NELSGDLCKEFGTTPKEPPEPTPLDELAEATFQDFPSG---ALGROSSLLRLISSKNP	238	
QY	195	QMMLAAMVKSYLEKKGI-APKD--MVMSIMPTCKROSEADRDWFCVDADPTLRQLD	249	
DB	239	GGFTTVARKYLQTRMGLPSGRQDVLVALSNEPARLGSBADAKAFL-----DSMAQRY	294	
QY	250	HYTTVEIGNIFKEKGINIAELPEDEMDNPMGVSGAGVLTFTTGYME	298	
DB	295	ASIVGVDSLASASASG-----MAGAGAAAGAMIDAGALEE	330	
RESULT 12				
ALR_STRCO		STANDARD;	PRT;	391 AA.
ID	ALR_STRCO			
AC	086786;			
DT	30-MAY-2000 (Rel. 39, Created)			
DT	30-MAY-2000 (Rel. 39, Last sequence update)			
DT	15-JUN-2002 (Rel. 41, Last annotation update)			
DE	Alanine racemase (EC 5.1.1.1).			
GN	ALR OR SC04745 OR SC664.23.			
OS	Streptomyces coelicolor.			
OC	Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;			
OC	Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomycetes.			
OX	NCBI_TaxID=1902;			
RP	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN-A3(2) / M145;			
RX	MEDLINE-21996410; PubMed-12000953;			
RA	Bentley S.D., Chater K.F., Cerdano-Tarraga A.-M., Challis G.L.,			
RA	Thompson N.R., James K.D., Harris D.E., Quail M.A., Kleser H.,			
RA	Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,			
RA	Croft A., Fraser A., Goble A., Hildalgo J., Hornsby T., Howarth S.,			
RA	Huang C.-H., Kleser T., Latke L., Murphy L., Oliver K., O'Neill S.,			
RA	Rabbinowitsch E., Rajandream M.A., Rutherford K., Rutter S.,			
RA	Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,			
RA	Warren T., Wietzorek A., Woodward J., Barrell B.G., Parkhill J.,			
RA	Hopwood D.A.;			
RT	"Complete genome sequence of the model actinomycete Streptomyces			
RT	coelicolor A3(2).";			
RL	Nature 417:141-147(2002).			
CC	-I- FUNCTION: Provides the D-alanine required for cell wall			
CC	biosynthesis (By similarity).			
CC	-I- CATALYTIC ACTIVITY: L-alanine = D-alanine.			

CC -1- COFACTOR: Pyridoxal phosphate (By similarity).  
 CC -1- PATHWAY: Along with D-alanine--D-alanine ligase, it makes up the  
 CC D-alanine branch of the peptidoglycan biosynthetic route.  
 CC -1- SIMILARITY: BELONGS TO THE ALANINE RACEMASE FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: AL031317; CAA20401.1; -  
 CC HSP: P10724; IBD0.  
 CC InterPro: IPR000821; Ala\_racemase.  
 CC Pfam: PF00842; Ala\_racemase.1.  
 CC PRINTS: PR00992; ALARACEMASE.  
 CC TIGRPFAMS: TIGR00492; alt; 1.  
 CC PROSITE: PS00395; ALANINE\_RACEMASE; 1.  
 CC Isomerase: Pyridoxal phosphate; Cell wall; Peptidoglycan synthesis;  
 KM Complete proteome.  
 FT ACT\_SITE 46 46 CATALYTIC BASE SPECIFIC TO D-ALANINE (BY  
 FT SIMILARITY).  
 FT ACT\_SITE 283 283 CATALYTIC BASE SPECIFIC TO L-ALANINE (BY  
 FT SIMILARITY).  
 FT BINDING 46 46 PYRIDOXAL PHOSPHATE (BY SIMILARITY).  
 SQ SEQUENCE 391 AA; 41199 MW; 24754AC1385DCA1B CRC64;  
 Query Match 4.28; Score 107.5; DB 1; Length 391;  
 Best Local Similarity 24.18; Pred. No. 2;  
 Matches 100; Conservative 50; Mismatches 130; Indels 135; Gaps 24;  
 Oy 5 VLKPCAIVSIRGSSCARQVAPRAPIAASVVALATLAEAPRLGVACAAAPAAEAP 64  
 Db 13 VLBARFIDIALRANVRALRERAPCA-----LMAVVKADATGCGATPCARAAVAGAT 67  
 Oy 65 ---LSHVQALA-ELAKP-KDD-----PTRKHV--CYQVAPARVAIETLG 104  
 Db 68 WLGTAPFOEALRAAEFGPDDVRIMCMLMTPGGRWREAVRLDVSAMAMEEVTG 127  
 Oy 105 LAGATTPKO-----LAEGRLGFEVFTLGADLTIEEGSELLHRLTEHLEAH 156  
 Db 128 AAAAACPAPARVOLKADTGLRGCGCPGAD--WERLYGCAALRAEEGLRLTGLMSHAC- 184  
 Oy 157 PHSDPELPMTSCPCGMIAMLEKSYPLIPYSSCKSPQMLAAVKSVAEKKGAPK- 215  
 Db 185 --ADP-----GHSI-----AAQTLRRREMTAV-AEQRGLAFEV 216  
 Oy 216 -----DMV-----MVSIMPTCKRQSEAD--RDMFCVDADPTL-RQL- 248  
 Db 217 RHIANSPATLTLDPADHDLVRPGCIAMGVSPSEIGTPADFGRLPVTMLAASLALVQVP 276  
 Oy 249 -----DHVIT-----VELG---NIFKRGINLAELPGECENDPMG----- 281  
 Db 277 GGHGVSYGHHTTTPGETTGLVPLGYADGIPRIASSGPLYVCKMTVAAGRIAMDQFV 336  
 Oy 282 -VGS-----GA-CVLFG-----TTGCVMEALRTAYELFT--GFFPLRLTSE 320  
 Db 337 DLGGDRPEGAELVLFPGDRGEPTADMAQAAGTATLAEIVTIGSNVPRVYVNE 391  
 RESULT 13  
 FAST CANAL STANDARD: PRT; 1885 AA.  
 AC P43098;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Fatty acid synthase subunit alpha (EC 2.3.1.86) [Includes: Acyl  
 DE carrier; 3-oxoacyl-[acyl-carrier protein] reductase (EC 1.1.1.100)  
 DE (Beta-ketoacyl reductase); 3-oxoacyl-[acyl-carrier protein] synthase  
 DE (EC 2.3.1.41) (Beta-ketoacyl synthase)].

GN FAS2.  
 OS Candida albicans (Yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; mitosporic Saccharomycetales; Candida.  
 ON NCBI\_TaxID=5476;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=4918;  
 RX MEDLINE=95255657; PubMed=737507;  
 RA Southern S.B., Cihlar R.L.;  
 RT "Analysis and expression of the Candida albicans FAS2 gene.";  
 RL Gene 156:133-138(1995).  
 CC -1- FUNCTION: FATTY ACID SYNTHETASE CATALYZES THE FORMATION OF  
 CC THE ALPHA SUBUNIT CONTAINS DOMAINS FOR: ACYL CARRIER PROTEIN,  
 CC 3-OXOACYL-[ACYL-CARRIER PROTEIN] REDUCTASE, AND 3-OXOACYL-[ACYL-  
 CC CARRIER-PROTEIN] SYNTHASE.  
 CC -1- CATALYTIC ACTIVITY: Acetyl-CoA + N malonyl-CoA + 2N NADPH = 2  
 CC long-chain fatty acid + (N+1) CoA + N CO(2) + 2N NADP(+).  
 CC -1- CATALYTIC ACTIVITY: Acyl-[acyl-carrier protein] + malonyl-[acyl-  
 CC carrier protein] = 3-oxoacyl-[acyl-carrier protein] + CO(2) +  
 CC [acyl-carrier protein].  
 CC -1- CATALYTIC ACTIVITY: (3R)-3-hydroxyacyl-[acyl-carrier protein] +  
 CC NADP(+) = 3-oxoacyl-[acyl-carrier protein] + NADPH.  
 CC -1- SUBUNIT: [Alpha(6)beta(6)] hexamers of two multifunctional  
 CC subunits (alpha and beta).  
 CC -1- SIMILARITY: TO THE FATTY ACID SYNTHETASE, SUBUNIT ALPHA FROM  
 CC OTHER FUNGI.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: L29063; AAA34345.1; -  
 CC InterPro: IPR002583; ACPS.  
 CC InterPro: IPR000794; Ketoacyl-synt.  
 CC InterPro: IPR004568; Pantethe\_ttn.  
 CC InterPro: IPR003880; Pantethe\_attach.  
 CC Pfam: PF00109; ketoacyl-synt; 1.  
 CC Pfam: PF01648; ketoacyl-synt; 1.  
 CC DR Pfam: PF02801; ketoacyl-synt\_C; 1.  
 CC DR ProDom: PD004282; ACPS; 1.  
 CC DR TIGRPFAMS: TIGR00556; pantethn\_ttn; 1.  
 CC DR PROSITE: PS00012; PHOSPHOPANTETHEINE; 1.  
 CC DR PROSITE: PS00606; B\_KETOACYL-SYNTHASE; 1.  
 CC KM Fatty acid biosynthesis; Multifunctional enzyme; Oxidoreductase;  
 KM transferase; NADP; Phosphopantetheine.  
 FT DOMAIN 1 2 ACYL CARRIER (ACP).  
 FT DOMAIN 1 2 BETA-KETOACYL REDUCTASE.  
 FT DOMAIN 1 2 BETA-KETOACYL SYNTHASE.  
 FT BINDING 181 181 PHOSPHOPANTETHEINE (BY SIMILARITY).  
 FT ACT\_SITE 1304 1304 BETA-KETOACYL SYNTHASE (BY SIMILARITY).  
 SQ SEQUENCE 1885 AA; 207588 MW; 4835D57F362372E0 CRC64;  
 Query Match 4.18; Score 106; DB 1; Length 1885;  
 Best Local Similarity 21.67; Pred. No. 20;  
 Matches 63; Conservative 36; Mismatches 111; Indels 82; Gaps 13;  
 Oy 20 RARQVAPRAPDLASTVVALATLAEAPRLGVACAAAPAAEAPLSHVQALAEALAKPK 79  
 Db 104 KOESTPSAPAAATPPRAAA--APF-----PAPAPASAGPVSI-----P 142  
 Oy 80 DDPTRKHVCQVAPARVAIAETLGLAGATTPKQALAEGLRGFDEVFDTLFGADLTIM 139  
 Db 143 DEPKANLLIHVLAQK--LKKPLDAVPMKAIKDLVNGKSTV----- 183  
 Oy 140 EEGSELLHRLTEHLEAHPHSDPELPK-----FTSCCPGMIAMLEKSYPLIPYSSCKS 193  
 Db 184 --ONEITGIDGKRGSTPEKPEEDTPIELAEQFODSFSC--DLGKSTSLIGRLMSSKM 238

Query Match	4.1%	Score 105.5	DB 1	Length 3491
Best Local Similarity	21.6%	Freq. No. 46		
Matches 117	Conservative	60	Mismatches 213	Indels 151
Gaps				25
OY	10	AAVSIRSSCARAPVAPRAPAASTVRALATLEAP	---RRIGNVCAAAAPAAEAPL	65
DB	405	ASTGTRHRTCCERLLA	-----VNGETALAGREDAATRTRELGLDSVLAQAARAK	456
OY	66	SHVOAL--AELAKPKDPTFRKHVCQVAVPAVVAIAETLC	-----	104
DB	457	--VSAALGREVNIALYDHPPRPALAELAACTEVAQRETRARTNEAPGEVAVVAMAC		514
OY	105	-LAPGATTPKQ---LALGLRL	-----GFDEYFDLTFGADLTIMEEGSGLHRLTEH	152
DB	515	RLPGGVSTPEEFWELLSGSRDAVAGLPTDRQMD--LDSLTFHDDPT--RSGT	-----	561
OY	153	LEAPHSDEPLPMFTSCCPGMIAMLEKRYPLPIPVSSCKSPOMLAAVKSSTYLAKEKGI		212
DB	562	--AHORGGGFTLEATFADPAFFGM-----SPEEALAVDPQORLMLELSNEVLERAGI		611
OY	213	APKVMVMSIMPCATRKQSEADRFECVDADPTLRDLQDVITTVELGATFKFERKGINIAELP		272
DB	612	PPSTL-----QASP-----TGVYFGLLPQETGRPALAE--		638
OY	273	EGEWDNPWGVSAGGVLEFGTTGGVMEALRTAYELEFTGTPPLPRLSLSEV	-----	321
DB	639	-----GGEGVEGLMTGTTSV--ASGRITAYTL--GLECPAISVDPAACSSSLVAVHILA		687
OY	322	-----RGMDGKET--NITWVAAPSGKFEELLKHHRAAAAEAAHGTPOPLAMDGAGFTS		375
DB	688	QSLRRGESSSLAMAGGYVMPPTPGMLVDFSRNSSLAPDGRKATSA-----GANGGGM		740
OY	376	EDRGGGTLIRVAVANGLGNNAKLLITTKQAGEAKYDEVEL--NACPGVCGVGGOPTSTOKA		434
DB	741	AEGGAKMLLE--RLSDARRNGHPIVLAVLGR7AVNSGASNGLSAP-----NGRAQVRVIOQA		795
OY	435	ITQR--ROALYLNDEKSTLRSHEHNSIRELYDTY--LGEP--LGHKAHELLTHHYA		487

Db 796 LAESGLPADIDAVEAHGTCTRLGDEPTEARALFEAYGRDREQLHLGYSKSNLGHQAAA 855  
 QY 488 G 488  
 Db 856 G 856

RESULT 15  
 ODP2\_RHIME  
 ID ODP2\_RHIME STANDARD; PRT; 447 AA.

AC Q9R9N3;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Dihydrolipoamide acetyltransferase component of pyruvate dehydrogenase complex (EC 2.3.1.12) (E2).  
 GN PDHC OR PDHB OR K01447 OR SMC01032.  
 OS Rhizobium meliloti (Sinorhizobium meliloti).  
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group; Rhizobiaceae; Sinorhizobium.  
 OX NCBI\_Taxid=382;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=RCR2011 / S047;  
 RX MEDLINE=20255031; PubMed=10796014;  
 RA Cabanes D., Boistard P., Batut J.;  
 RT "Symbiotic induction of pyruvate dehydrogenase genes from Sinorhizobium meliloti."  
 RT Mol. Plant Microbe Interact. 13:483-493(2000).  
 RL [2]  
 RN SEQUENCE FROM N.A.  
 RP STRAIN=1021;  
 RC MEDLINE=21396507; PubMed=11481430;  
 RA Capela D., Barloy-Hubler F., Gouzy J., Bothe G., Ampe F., Batut J., Boistard P., Becker A., Boutry M., Cadieu E., Dreano S., Gloux S., Godrie T., Goffeau A., Kahn D., Kiss E., Lelure V., Masuy D., Pohl T., Portetault D., Puehler A., Purnelle B., Ramsperger U., Renard C., Thebaud P., Vandenbol M., Weidner S., Galibert F.;  
 RT "Analysis of the chromosome sequence of the legume symbiont Sinorhizobium meliloti strain 1021."  
 RT Proc. Natl. Acad. Sci. U.S.A. 98:9877-9882(2001).  
 RL  
 CC - FUNCTION: THE PYRUVATE DEHYDROGENASE COMPLEX CATALYZES THE OVERALL CONVERSION OF PYRUVATE TO ACETYL-COA & CO(2). IT CONTAINS MULTIPLE COPIES OF THREE ENZYMAIC COMPONENTS: PYRUVATE DEHYDROGENASE (E1), DIHYDROLIPOAMIDE ACETYLTRANSFERASE (E2) & LIPOAMIDE DEHYDROGENASE (E3) (BY SIMILARITY).  
 CC - CATALYTIC ACTIVITY: Acetyl-CoA + dihydrolipoamide - CoA + S-acetyldihydrolipoamide.  
 CC - COFACTOR: THE E2 COMPONENT CONTAINS ONE COVALENTLY-BOUND LIPOYL COFACTOR (BY SIMILARITY).  
 CC - SUBUNIT: FORMS A 24-POLYPEPTIDE STRUCTURAL CORE WITH OCTAHEDRAL SYMMETRY (BY SIMILARITY).  
 CC - SIMILARITY: BELONGS TO THE 2-OXOACID DEHYDROGENASE FAMILY.  
 CC - SIMILARITY: CONTAINS 1 LIPOYL-BINDING DOMAIN.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; AF190792; AAF04589.1; -  
 CC EMBL; AL591787; CAC46026.1; -  
 CC HSSP; P07016; 1BBL.  
 CC InterPro; IPR001078; 2oxoacid\_dh.  
 CC InterPro; IPR000089; Biotin\_lipoyl.  
 CC InterPro; IPR004167; E3 binding.  
 CC InterPro; IPR003016; lipoyl.  
 CC Pfam; PF00198; 2-oxoacid\_dh; 1.  
 CC Pfam; PF00364; biotin\_lipoyl; 1.

DR Pfam; PF02817; e3\_binding; 1.  
 DR ProDom; PD00115; 2oxoacid\_dh; 1.  
 DR POSITE; PS00189; LIPOYL; 1.  
 KW Glycyls; Transferase; Acyltransferase; lipoyl; Complete proteome.  
 FT BINDING 43 43 LIPOYL (BY SIMILARITY).  
 FT ACT\_SITE 420 420 POTENTIAL.  
 SQ SEQUENCE 447 AA; 46140 MW; 48B1CAG4E2PDC2AC CRC64;

Query Match 4.1%; Score 105; DB 1; Length 447;  
 Best Local Similarity 19.5%; Pred. No. 3.9;  
 Matches 99; Conservative 57; Mismatches 161; Indels 190; Gaps 23;

QY 11 AVSIRSSCRARQVAPAPAPLAATVAVATLEAPAPRLGNVACAAAPAAEAPL----- 65  
 Db 85 ATAKGNGAGAVP--APRKETAETAPAAAPAP--AAPAAQAAAPSPAPADGEGK 139  
 QY 66 ----SHVOALAE-----LAKKDPTRKRVVQVAPAVRAIAETLGIAPATTPKOL 115  
 Db 140 RIFSSPLARRLAEGIDLSAIGSGPHGRVVKVETAVAGGAAPAPAPAPAPATL 199  
 QY 116 AECRLRRGFEVDFLEADLTMEGS-----ELRLTEHLEAHPSPDEPLP 164  
 Db 200 AKGMS-----DAVLKLFEPGSYELVPHDGMKRTIAKRLVESKQTIPF----- 243  
 QY 165 MFTSCCPGWTAMLEKSTPDLIPYVS--SCSPQMA--LAAMVKSILAERKQ-----TAPKDM 217  
 Db 244 -----YVSVCEDLALMALRAQLMAAPAPKPDGPVYKLSVNDM 281  
 QY 218 VMSVIMPCTRKQSEADRDMECVADPTLRDLHVITTVELGNIKFERGINLAEELPGEMD 277  
 Db 282 VIKALALALRDVVDANVSW-----TDONMKVHKH----- 310  
 QY 278 NPMGVSGAGVLEGTGGVNEALRTAYLEFTGTPPLRLSLSEVRGMDGKETNITWPA 337  
 Db 311 -----ADVGVANVINGGLT-----TPI-----VRQAEKLSLSA 338  
 QY 338 PGSKFEELKRRAAARAAAHAGTPOPLANDGAGFTSEDRGCTLRVAVAN--GLGNK 336  
 Db 339 ISNEMKDILK--RAKERK-----LKEPEYQGGTT--AVSNMGMGVYK 376  
 QY 397 KLITKMGAGEAKYDVEYIMAPAG--CVGSGGQPRSTDAIKQKRAALYNLDEKSTLR 453  
 Db 377 -----DFAAVNPPIATITAVAG-----EDRVVVRNKMVAVIANV--MYTTL 417  
 QY 454 RSHE-----NPSIRELYDTYLGEPLG 474  
 Db 418 TDHRCVDGALGAEELAAFKRYIENPMG 444

Search completed: June 3, 2003, 16:16:21  
 Job time : 25 secs